

TOWARD DISTANT READING OF THE NEW TESTAMENT COMMENTARY:
EXPLORING BIBLIOGRAPHICAL AND STRUCTURED TEXTUAL DATA

by

RICHARD MANLY ADAMS, JR.

STEVEN MACCALL, COMMITTEE CHAIR
EMMA ANNETTE WILSON
DAN ALBERTSON

A THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Library and Information Studies
in the School of Library and Information Studies of
The University of Alabama

TUSCALOOSA, ALABAMA

2017

Copyright Richard Manly Adams, Jr. 2017
ALL RIGHTS RESERVED

ABSTRACT

This study argues for distant reading techniques as a new form of understanding the genre of the New Testament commentary. Though it is a dominant form of New Testament scholarship, the commentary is a difficult genre to understand, due to the size of the corpus and the close reading techniques common in Biblical studies. The scholar who wants to understand the genre, therefore, is forced to choose representative samples, a subjective process that distorts understanding of the genre. This study introduces a reading technique that looks for patterns of development over this large corpus, distant reading that will help scholars understand the changes in the commentary over time and help identify places where close reading is warranted. To introduce an alternative form of reading this study demonstrates two modes of reading the genre. The first is a mining of the bibliographic metadata to demonstrate trends in the publication patterns and sizes of New Testament commentaries. The second is to structure commentaries based on their treatment of individual chapters and verses of New Testament texts, as a way of charting the priorities of given commentators over time. The study concludes with an invitation to librarians to work to better prepare the corpus of the New Testament commentary for distance reading.

ACKNOWLEDGEMENTS

This project was the culmination of my wonderful experience at the University of Alabama School of Library and Information Studies. I would like to thank the work of the faculty there, most notably the members of my committee, Professor Steven MacCall, Professor Emma Wilson, and Professor Dan Albertson (now of SUNY-Buffalo). Other faculty members who helped along the way were Stephen Yates and John Burgess. I would also like to thank Beth Riggs for her magisterial coordination of that program and her helping me navigate the administrative side of things. My colleagues at the Pitts Theology Library, Emory University, were helpful conversation partners. I would also like to thank Daniel Moody for some early tagging he did on commentaries on 1 Corinthians as a test case for chapter 3. As always, the highest praise and appreciation should go to my family, who put up with me doing yet another degree. I promise, no more...

CONTENTS

ABSTRACT.....	ii
ACKNOWLEDGEMENTS.....	iii
LIST OF FIGURES.....	v
CHAPTER 1. DISTANT READING AND THE NEW TESTAMENT COMMENTARY.....	1
CHAPTER 2. READING THE BIBLIOGRAPHIC METADATA OF COMMENTARIES.....	21
CHAPTER 3. READING COMMENTARIES AS STRUCTURED TEXTUAL DATA.....	50
CHAPTER 4. THE PROMISE OF DISTANT READING IN BIBLICAL STUDIES.....	78
REFERENCES	82
APPENDICES.	86

LIST OF FIGURES

1. Publication Levels of Commentaries on all Books of the New Testament.....	31
2. Publication Levels of Commentaries on the Four Canonical Gospels (1901-2017).....	32
3. Publication Levels of Commentaries on Matthew, Acts, 1 Corinthians, Hebrews (1901-2017).....	32
4. Publication Levels of Commentaries on John and Revelation (1901-2017).....	33
5. Publication Levels of Commentaries on Each of the Four Canonical Gospels (1901-2017)..	34
6. Publication Levels of Commentaries on Mark (1801-1900).....	34
7. Publication Levels of Commentaries on Each of the Thirteen Pauline Letters (1901-2017)..	35
8. Publication Levels of Commentaries on Galatians and Philippians (1941-1970).....	35
9. Publication Levels of Commentaries on the Disputed Pauline Letters (1941-2017).....	36
10. Growth of OT and NT Studies Compared with Other LCSH (1901-2017).....	37
11. Number of Publications by Decade for Given Subject Headings (1901-2010).....	38
12. Publication Levels for Works with Given Subject Headings (1901-2010).....	38
13. Average Number of Pages for Commentaries on Revelation Published within a Given Year (1950-2016).....	41
14. Average Number of Pages for Commentaries on Mark Published within a Given Year (1950-2015).....	42
15. Average Number of Pages for Commentaries on Mark Published within a Given Year (1901-1950).....	43
16. Commentaries on Acts Published by Year (1901-2015).....	43
17. Comparison of the Relative Treatment of Each Chapter in 1 Corinthians, Fee and Calvin....	52

18. Comparison of the Relative Treatment of Each Verse in 1 Corinthians, Fee and Calvin.....	53
19. Comparison of the Relative Treatment of Each Verse in 1 Corinthians 11, Fee and Calvin..	53
20. The 8 Commentaries in the Corpus Studied.....	65
21. A Word Cloud of the Content of the 8 Commentaries.....	66
22. Standard Deviations of Verse Treatments.....	67
23. Percentage of Words Written on 3:12, From Earliest to Most Recent.....	69
24. Average Word Counts Per Verse over the Entire Corpus.....	71
25. Number of Articles Indexed by Verse of James in the ATLA Religion Database.....	73

CHAPTER 1. DISTANT READING AND THE NEW TESTAMENT COMMENTARY

“Which of us can read all that he has written?” – St. Jerome¹

“The digital revolution, for all its wonders, has not penetrated the core activity of literary studies, which, despite numerous revolutions of a more epistemological nature, remains mostly concerned with the interpretive analysis of written cultural artifacts.” – Stephen Ramsay²

This exploratory study investigates a new path by which studies of the genre of Biblical commentary might proceed. It does so by introducing the method of distant reading and topic modeling by computational analysis to Biblical studies as a way of better understanding this important aspect of the history of Biblical interpretation. The underlying assumption of the study is that the close reading practices of the field of Biblical scholarship form a major hindrance to systematic analysis of the development of this large corpus of literature. I introduce distant reading techniques, developed outside of Biblical studies and within information studies and the digital humanities, as method to help scholars better understand how the genre has changed over time, and therefore potentially re-shape the conversation about the commentary.

In this study, I introduce distant reading methods through computational analysis as a way to read the genre of New Testament commentary.³ After introducing the method, I identify

¹ Jerome, “Letter to Pammachius and Oceanus,” 8 (*Nicene and Post-Nicene Fathers* 6).

² Stephen Ramsay, *Reading Machines: Toward an Algorithmic Criticism* (University of Illinois Press, 2011), 2.

³ The computational analysis of the study will focus exclusively on commentaries written about books of the New Testament. This is simply a choice of convenience, and the conclusions drawn

two new data sets as potential areas to apply distance reading techniques: bibliographic metadata and the format of the commentary itself as structured textual data. In the second and third chapters of the study, I demonstrate how these two data sets can be manipulated to reveal information about the development of the genre through distance analysis. It is my intention that this study function more as an invitation than a full demonstration. That is, I write to encourage scholars who work with other large corpora of texts to see the power of this hermeneutical shift from close analysis to pattern detection over time and texts, and this encouragement comes in the form of identifying possible areas of analysis. The power of the distant reading method is the promise of bringing together disparate texts within a single tradition and to read textual traditions anew, overcoming the canonical inertias that plague all areas of the humanities.

The central argument to the study is that understanding the nature of the genre of Biblical commentary, or the subset of New Testament commentary demands critical analysis of a large portion of the enormous corpus. To properly analyze the genre, interpreters must shift their understanding of the corpus and their method of reading. Rather than reading closely for content a few representative samples, interpreters should first read from a distance the bibliographic metadata of the commentary genre and the textual structure of large numbers of commentaries. This study demonstrates that the practice of distant reading through computational analysis is well-suited to the study of the commentary as genre.

The study is organized into four chapters. The first is an introduction to distant reading and argument for its need in understanding the commentary tradition. In the second and third chapters, I introduce two data sets appropriate for distant reading the commentary tradition, and

about the development of this set of literature could likewise be applied to a study of commentaries on the Hebrew Bible.

in each I offer a brief demonstration of applying distant reading techniques to the respective data sets. In chapter 2, I introduce bibliographic metadata as a resource for scholars to analyze how the publication rate and size of the New Testament commentary has changed over time. In chapter 3, I introduce the verse-by-verse form of the commentary as an inherently structured textual data set that is connected to a commentary's content, and I demonstrate the promise of analyzing changes in form as a clue to developments in the approach to particular New Testament texts. In the final chapter of the study, I suggest ways forward for librarians and subject matter experts working together to take advantage of distant reading techniques.

In this introduction, I lay out the reasons for, methodology of, and benefits from this exploratory study. In a first section, I identify the nature of the study of New Testament commentary that defines the need for distant reading, including the challenge presented by the large corpus of New Testament commentaries and the tradition of close reading in New Testament studies. I then introduce the method of distant reading, an approach that embraces a large corpus by looking for patterns of change over time. With the background of the issues in reading New Testament commentaries established and the introduction of the theoretical framework of distant reading, I turn to identify specific research questions about distant reading and the commentary tradition and the specific methodical steps this study takes to begin to address these questions. I close with a consideration of the impact of this study on the field of New Testament studies. This opening chapter explains the need for distant reading; the subsequent chapters will explore two aspects of the corpus of New Testament commentaries that scholars could distant read to help trace the development of this genre.

Definitions

Before proceeding, it may be helpful to offer working definitions for the two key terms at the center of this study. The first is “commentary.”⁴ By this term, I mean written analysis of the Biblical text that offers discrete annotations, organized by the chapters and verses of the Biblical text itself. The content of this analysis may be philological, historical, literary, or theological, but a commentary is identified by the organization of those comments to follow the versification of the Biblical text; the designation “commentary” is a distinction of form, not content. In this study, I limit my scope to the subset of New Testament commentaries, such treatments focused on the 27 books of the New Testament.

This study introduces distant reading as a way of understanding the genre commentary, and so a second key definition is “distant reading.” Though more discussion of this term follows below, here I define distant reading as analysis of a large corpus of written texts (larger than a single work) that focuses on development across texts or over time. That is, in contrast to close reading which takes as its aim understand the idiosyncrasy of a single interpretation, distant reading focuses on understanding patterns across a genre, time period, or other cluster of analysis.

⁴ For a consideration of the definition of this genre, see Lars Hartman, “A Commentary: A Communication about a Communication,” *Novum Testamentum* 51, no. 4 (2009): 389–400, esp. 389-91; Markus N. A. Bockmuehl, “The Dead Sea Scrolls and the Origins of Biblical Commentary,” pp. 3-29 in *Text, Thought, and Practice in Qumran and Early Christianity: Proceedings of the Ninth International Symposium of the Orion Center for the Study of the Dead Sea Scrolls and Associated Literature, Jointly Sponsored by the Hebrew University Center for the Study of Christianity, 11-13 January, 2004* (Leiden: Brill, 2009), esp. 4-6. I adopt here a variation on the loose definition from Bockmuehl: “Works consisting primarily of sequential, expository annotation of identified texts that are themselves distinguished from the comments and reproduced intact, whether partially or continuously” (“Dead Sea Scrolls,” 4). For consideration of the roots of the contemporary definition of commentary, see Marina Stojanović, “On the Genre of Commentary in the Biblical Exegesis,” *Philotheos* 15 (2015): 70.

Research Objectives

This study sets as its objectives the following:

1. Outline the need for and introduction the methods of distant reading to students of the humanities and argue effectively for this as a method for reading the history of New Testament commentary, and Biblical interpretation in general;
2. Explore computational analysis on the publication metadata of New Testament commentaries and describe a method by which such analysis could overcome the “canon” problem of traditional methods of studying the genre;
3. Identify the form of the New Testament commentary as a consistent element of the genre and outline a method of distant reading by parsing relative verse and chapter treatments as a way of tracing the development of the genre of New Testament commentary, with specific focus on the letter of James.

Background

The Need for Distant Reading in Biblical Studies

Interpreters of the commentary tradition face an enormous corpus of literature to analyze, a corpus that has developed over millennia and grows by hundreds of publications each year. This problem of a large corpus is exacerbated by the practices of close reading, practices that define much of the history of Biblical interpretation. Interpreters in fields outside of Biblical studies have identified these twin problems in considering other corpora of literature. In response, some have developed distant reading techniques, in which the large size of a corpus is embraced, not as the object of the synchronic practice of close reading of individual samples, but rather for diachronic analysis of trends, patterns, and developments over the entire corpus.

The digital humanities scholar would argue that in order to understand the development of New Testament commentaries, Biblical scholars need to move from dependence upon close reading to being open to distant reading techniques. To invoke the well-known phrase of Franco

Moretti, in order to understand the development of New Testament commentaries, we need to “learn how *not* to read them.”⁵

Commentaries in Biblical Scholarship

The commentary has been the dominant form of Biblical scholarship for thousands of years. David Clines calls “commentary-writing the quintessential form of biblical scholarship.”⁶ While the hermeneutical assumptions behind, methods of, and participants in Biblical scholarship have shifted generation to generation, the genre of the verse by verse commentary has remained the primary mode of interpretation, in both Jewish forms, dating at least as early as the Dead Sea Scrolls (ca. 1st C BCE), and Christian forms, dating at least to Hippolytus of Rome (early 3rd C CE). As Brevard Childs notes, “From the earliest period, first in Judaism and later in Christianity, the commentary provided the classic form by which Scripture was both studied and interpreted. Even in moments when a traditional understanding of the Bible was being challenged or a new model proposed, the genre of the commentary continued as a vehicle of change.”⁷ While the ubiquity of the commentary as a form of interpretation is not unique to the study of the Bible, it is a defining characteristic of the tradition of Biblical scholarship.⁸ Drawing

⁵ Franco Moretti, “Conjectures on World Literature,” *New Left Review*, II, no. 1 (2000): 57.

⁶ David J. A. Clines, “Writing a Job Commentary,” in *The Genre of Biblical Commentary* (Eugene, OR: Wipf and Stock, 2015), 38.

⁷ Brevard S. Childs, “The Genre of the Biblical Commentary as Problem and Challenge,” pp. 185-92 in *Tehillah le-Moshe: Biblical and Judaic Studies in Honor of Moshe Greenberg* (ed. Mordechai Cogan, Barry L. Eichler, and Jeffrey H. Tigay; Winona Lake, IN: Eisenbrauns, 1997), 185.

⁸ Childs seems to understand why the genre was popular early in the church’s history, but he notes that, “An issue more difficult to explain is why the form of the commentary has continued ‘alive and well,’ even when many of the traditional assumptions regarding the nature of the Bible are no longer shared by the modern academic culture” (“Genre,” 186).

upon the developed tradition of classical and Jewish commentaries, the early Christians quickly took to the form of verse by verse comments on sacred texts.⁹

Despite the ubiquity of the Biblical commentary as a form of scholarship, there has been surprisingly little study of the genre itself.¹⁰ The rising interest in the history of the Bible's reception and influence has led some to ask questions about why the early church turned to the genre of commentary or what makes this genre so popular amongst Biblical scholars.¹¹ Other than a few studies, though, analysis of the development of the genre of commentary is a real *desideratum* of Biblical scholarship. Even a recent collection that purportedly focuses on this

⁹ The commentary as a line by line analysis of authoritative texts is a development of the Hellenistic period, focused primarily around the works of Homer. For discussion, see L.D. Reynolds and Wilson, N. D., *Scribes and Scholars: A Guide to the Transmission of Greek and Latin Literature*, 3rd ed. (Oxford: Clarendon Press, 1991), 9-15. Jewish commentaries on Scripture, known as *pesharim*, were discovered amongst the texts at Qumran, and pre-date Christian commentaries. It is not clear whether there is a direct connection between these *pesharim* and what came to be Christian commentaries. For discussion, see Bockmuehl, "Dead Sea Scrolls," esp. 14-19.

¹⁰ Frequently-cited studies of the development of the genre of Biblical commentary are found in the *Cambridge History of the Bible*. See, in particular, B. Smalley, "The Bible in the Medieval Schools," pp. 197-220 in *Cambridge History of the Bible 2* (Cambridge: Cambridge University Press, 1969). There have been a few studies that look at the genre, but none systematically. For examples, see Bernhard W. Anderson, "The Problem and Promise of Commentary," *Interpretation* 36, no. 4 (1982): 341-55; John Barton, "Biblical Commentaries," *Epworth Review* 24, no. 3 (1997): 35-44; Adele Berlin, "On Bible Translations and Commentaries," in *Bible Translation on the Threshold of the Twenty-First Century: Authority, Reception, Culture and Religion* (London: Sheffield Academic Press, 2002), 175-91; Richard J. Coggins, "A Future for the Commentary," pp. 163-75 in *Open Text: New Directions for Biblical Studies?* (London: SCM Press, 1993); Lars Hartman, "A Commentary: A Communication about a Communication," *Novum Testamentum* 51, no. 4 (2009): 389-400; Steven J. Koskie, "Seeking Comment: The Commentary and the Bible as Christian Scripture," *Journal of Theological Interpretation* 1, no. 2 (2007): 237-49; John Nolland, "The Purpose and Value of Commentaries," *Journal for the Study of the New Testament* 29, no. 3 (2007): 305-11; Stojanović, "On the Genre of Commentary."

¹¹ Childs offers four rationales for the dominance of the commentary: the early privileging of the Biblical text itself led to a distinction between text and interpretation; the early focus on the maintenance of the original order of the text created the need for a line by line analysis; the notion of an inspired text led to the focus on every word of the text; the connection of the text to the life of communities necessitated close analysis by trained experts (See Childs, "Genre," 185-86).

genre pays little attention to its development over time, and focuses instead on individual commentaries and commentary writers.¹² As Bernhard W. Anderson notes, “The biography of this genre, commentary, has not yet been written, at least to my knowledge.”¹³

I propose two major reasons that there has not been systematic study of the genre of Biblical commentary. The first is the size of the corpus to be analyzed. Ironically, it is likely the dominance of the commentary as a genre that is itself responsible for the challenge of studying it. With the production of hundreds of commentaries on the books of the Bible each year, it seems impossible to track the development of the genre; one simply cannot read all the commentaries produced on a given book of the Bible, let alone the entire Biblical canon.¹⁴ Scholars interested in mapping developments are forced to look at particular incidents or dominant figures so as to identify pivot points in the development of the genre. Consider, for example, the important essay of Brevard Childs, one attempt to show how the commentary has changed over time. Childs identifies three issues that have shaped the development of the modern commentary: the emergence of secular Biblical scholarship; the shift between popular and scholarly audiences; the diversification of hermeneutical approaches to the text. To establish each point, though, Childs must generalize from specific examples to what seems to be a dangerous degree. For example, he identifies a shift in Genesis scholarship from treating the text in its canonical order toward rearranging the text to match what scholars deem to be the chronological order of the development of the text. To mark this shift, though, he cites only one late 19th century

¹² William Yarchin, *The Genre of Biblical Commentary: Essays in Honor of John E. Hartley on the Occasion of His 75th Birthday*, ed. Timothy D. Finlay (Eugene, Oregon: Pickwick Publications, 2015).

¹³ Anderson, “Problem and Promise,” 343.

¹⁴ Consider, for example, the number of commentaries just for the gospel of John OCLC lists for 2008 (53), 2009(48), 2010 (53), 2011 (53), 2012 (37), and 2013 (33).

commentary (Dillman) as characteristic of the former position and two early 20th century commentaries (Gunkel and Procksch) as characteristic of the latter.¹⁵ It is simply not possible to read all commentaries, even of a narrow chronological era, and safely draw conclusions about the genre.

The second problem with studying the genre of Biblical commentary is closely related to the first: the entrenched practice of close reading. The dominance of verse by verse commentary is a reflection of the close reading practices of Biblical scholars. That is, the fact that the scholarly tradition operates largely through a form of secondary literature that consists of comments on individual verses of the Biblical text demonstrates how highly Biblical scholars value close reading of individual words, sentences, and units.¹⁶ This is true in the historical-critical mode of Biblical scholarship dominant from the Reformation through the 20th century, as well as more recent forms of interpretation that look for deeper meanings encoded within the text, what Stephen Best and Sharon Marcus identify as “symptomatic reading”: “A specific type [of interpretation] that took meaning to be hidden, repressed, deep, and in need of detection and disclosure by an interpreter.”¹⁷ For both the historical critic as well as the structuralist critic, the goal of interpretation is to find meaning from a given text. This meaning is not limited to the language of the text, but the text must be considered as part of a broader conversation. The language of the text is understood as a lens through which meaning, understood in diverse ways,

¹⁵ See Childs, “The Genre,” 187.

¹⁶ For a discussion of the connection between close reading and the genre of the commentary, see William Yarchin, “Introduction to the Volume,” in *The Genre of Biblical Commentary* (Eugene, OR: Wipf and Stock, 2015), 1–9.

¹⁷ Stephen Best and Sharon Marcus, “Surface Reading: An Introduction,” *Representations* 108, no. 1 (2009): 1. For a discussion of the dominance of the historical critical approach to the Bible, seeing the text as a reflection of the context(s) of its creation, see John H. Hayes and Carl R. Holladay, *Biblical Exegesis: A Beginner’s Handbook*, 3rd edition. (Louisville, KY: Westminster John Knox Press, 2007), 1-33.

is reflected/refracted. With these hermeneutical assumptions in place, the task of Biblical scholarship has expanded beyond philological analysis of the text itself to exploring a surplus meaning(s) encoded in the language of the text but dependent upon other domains, be they historical, cultural, theological, etc. The result, therefore, is the proliferation of massive tomes on these small texts. It is not uncommon to find pages of commentary on a single word of the Biblical text. There are certainly theological reasons behind the importance of close reading in the tradition of Biblical scholarship, as the traditional ascription of authority, at times unique authority, to the words of the text necessitates drawing as much meaning as possible from the given words. For a variety of reasons, Biblical scholars are in the habit of close analysis of small sets of texts. That is, by definition, the nature of Biblical scholarship.¹⁸

And so, readers are left with thousands of commentaries in the hands of scholars who are trained in the art of close analysis of a small set of texts. It is no surprise that we do not find detailed studies of the commentary tradition at large. Close reading habits on such a large corpus lead necessarily to the selection of a few commentaries as representative of entire periods of interpretation. This project introduces a new way forward.

¹⁸ Consider the fact that the Greek New Testament is itself only 138,020 words, roughly half the length of James Joyce's *Ulysses*. And yet, New Testament scholarship is printed in dozens, if not hundreds, of scholarly journals, accounts for entire departments in seminaries and religious studies departments. As an example of the extreme close reading of the New Testament, consider that Craig S. Keener recently published a commentary of 1,104 pages, which, in addition to introductory matters, covers his exegesis of only Acts 1:1-2:47 (Craig S. Keener, *Acts: An Exegetical Commentary: Introduction and 1:1-2:47* [Grand Rapids, MI: Baker Academic, 2012]).

Distant Reading Methods and the Digital Humanities

While New Testament scholarship, and the analysis of that scholarship, continues to be defined by its close reading techniques, an alternative form of “reading” has been developing in the field of information studies and the digital humanities. Rather than focusing on the selection and close analysis of a few representative texts, digital humanists have argued for broad analysis of large corpora of texts, searching for patterns and development across many texts, rather than idiosyncrasies found through careful analysis of a single text.¹⁹ Through the methods of distant reading over many texts, scholars are able to discern the shifting forms, patterns, and linguistic uses within a given corpus.

Modern use of computational tools to analyze large corpora of texts is a movement often connected with the seminal work of the Italian literary scholar Franco Moretti.²⁰ In his essay “Conjectures on World Literature,” Moretti identifies the central problem associated with the close reading practices in traditional humanities scholarship, a problem which is familiar to those who have studied the development of the Biblical commentary. Moretti notes that with large corpora of texts (he was working primarily on European novels of the 19th century), scholars are forced to focus on small portions of the corpus: “The trouble with close reading (in all of its

¹⁹ For examples of calls for distant analysis of large corpora of texts, see Gina L. Greco and Peter Shoemaker, “Intertextuality and Large Corpora: A Medievalist Approach,” *Computers and the Humanities* 27, no. 5/6 (1993): 349–55; Matthew G. Kirschenbaum, “The Remaking of Reading: Data Mining and the Digital Humanities” (2009); Timothy R. Tangherlini and Peter Leonard, “Trawling in the Sea of the Great Unread: Sub-Corpus Topic Modeling and Humanities Research,” *Poetics* 41, no. 6 (2013): 725–49.

²⁰ For a clear introduction to the contribution of Moretti to literary studies, see Simon Ryan, “The Location of Knowledge,” *International Journal of Technology, Knowledge & Society* 8, no. 6 (2012): 73–81. Moretti’s theoretical work is most readily-available in the collection of his essays *Distant Reading* (London ; New York: Verso, 2013) and the monograph Franco Moretti and Alberto Piazza, *Graphs, Maps, Trees: Abstract Models for Literary History* (London; New York: Verso, 2007).

incarnations, from the new criticism to deconstruction) is that it necessarily depends on an extremely small canon.”²¹ The result of close reading is what Moretti calls a “theological exercise”: scholars selecting a small set of texts for their close analysis, basing their selection on the assumption that these few selected are representative of a genre as a whole. The limitations of human reading abilities combined with large sets of texts not only leads to the problem of “canon” and the recurring treatment of the privileged few texts, it also renders impossible the task of drawing any conclusions about a genre of literature. As James E. Dobson summarizes, “It is precisely the failure of abstraction foreclosed by the specificity of the singular close reading that motivates Moretti’s desire for a distanced position capable of producing a systemic critique.”²²

Moretti creates this “distanced” position by embracing large corpora of texts, but doing so by shifting the mode of reading. Rather than close analysis familiar to literary scholars, Moretti calls for “distant reading,” wherein the scholar attempts to understand “systems” of literature by focusing on development across time in the aggregate. He identifies distance between the reader and the text, a product of a large corpus at the center of the analysis, to be “a condition of knowledge: it allows you to focus on units that are much smaller or much larger than the text: devices, themes, tropes—or genres and systems.”²³ Moretti’s work looks at small elements across texts so as to understand how that element functions within the system of the set of texts.

²¹ Moretti, “Conjectures,” 57.

²² James E. Dobson, “Can an Algorithm Be Disturbed? Machine Learning, Intrinsic Criticism, and the Digital Humanities,” *College Literature* 42, no. 4 (2015): 546.

²³ Moretti, “Conjectures,” 57.

Moretti's work, particularly his emphasis on "systems of literature," has generally been well-received, not only in literary studies, but in a number of fields that face the task of reading large sets of texts.²⁴ The incorporation of distant reading techniques has produced insights in areas of literature as diverse as the novels of Gertrude Stein,²⁵ thesauruses of the English language,²⁶ British newspapers,²⁷ and translations of European poetry.²⁸

These studies that adapt Moretti's theoretical model of reading demonstrate that the presence of digitized texts and computational tools do not simply allow interpreters to read more efficiently. That is, the promise is not that scholars can now perform close reading at a faster pace, and, therefore, consume more works.²⁹ Rather, the opportunity exists for new types of analysis; distant reading creates the opportunity for insight that close reading, even at an efficient pace, could not reveal. This opportunity arises because distant reading seeks out patterns and shifts, rather than idiosyncrasies. It is because of this recognition of the opportunity for new

²⁴ Moretti's distant reading has been met with some criticism, particularly by those criticizing his statistical analysis as a way of supporting conclusions he had already made about the literature he studies. For a strong critique of Moretti's method, see Katie Trumpener, "Critical Response I: Paratext and Genre System: A Response to Franco Moretti," *Critical Inquiry* 36, no. 1 (2009): 159–71.

²⁵ Tanya E. Clement, "'A Thing Not Beginning and Not Ending': Using Digital Tools to Distant-Read Gertrude Stein's *The Making of Americans*," *Lit Linguist Computing* 23, no. 3 (2008): 361–81.

²⁶ Marc Alexander et al., "Metaphor, Popular Science, and Semantic Tagging: Distant Reading with the Historical Thesaurus of English," *Digital Scholarship in the Humanities* 30 (2015): 16–27.

²⁷ Dallas Liddle, "Reflections on 20,000 Victorian Newspapers: 'Distant Reading' The Times Using The Times Digital Archive," *Journal of Victorian Culture (Routledge)* 17, no. 2 (2012): 230–37.

²⁸ Jacob Blakesley, "Examining Modern European Poet-Translators 'Distantly,'" *Translation & Literature* 25, no. 1 (2016): 10–27.

²⁹ Matthew L. Jockers notes that Moretti does discuss just how difficult it would be to master the 19th Century British novels he is studying: "Assuming that a dedicated scholar could find these novels and read one per day, it would take sixteen and a half years of close reading to get through them all" (*Macroanalysis: Digital Methods and Literary History* [Urbana: University of Illinois Press, 2013], 19).

ways of analyzing texts that Matthew Jockers prefers to speak of “macroanalysis” rather than “distant reading.” In Jockers’ envisioned new forms of “reading,” “the very object of analysis shifts from looking at the individual occurrences of a feature in context to looking at the trends and patterns of that feature aggregated over an entire corpus.”³⁰ “Reading,” traditionally understood, is a qualitative process. What Jockers envisions is more data- and visualization-driven.

One particularly promising application of Moretti’s distant reading for the study of the commentary genre is topic modeling. Moretti’s emphasis on understanding literary corpora as “systems” has led to the development of “topic modeling,” attempts to allow algorithms to detect clusters of content in large sets of text. Andrew Goldstone and Ted Underwood note that “the aim of topic modeling is to identify the thematic or rhetorical patterns that inform a collection of documents.”³¹ Topic modeling has been applied to fields ranging from URL link disambiguation,³² to constitutional law,³³ and more recently to literary studies.³⁴ Rather than relying on a given interpreter’s pre-conception of what a text is “about,” topic modeling invokes the text mining powers of computers to help define what texts are talking about, identifying alliances across corpora that might not be apparent to the interpreter who reads with his or her

³⁰ Jockers, *Macroanalysis*, 24.

³¹ Andrew Goldstone and Ted Underwood, “The Quiet Transformations of Literary Studies: What Thirteen Thousand Scholars Could Tell Us,” *New Literary History* 45, no. 3 (2014): 361. Note that Goldstone and Underwood are interested in mining scholarly literature in the form of published articles.

³² Bradley Skaggs and Lise Getoor, “Topic Modeling for Wikipedia Link Disambiguation,” *ACM Transactions on Information Systems* 32, no. 3 (2014): 10–10:24.

³³ Daniel Taylor Young, “How Do You Measure a Constitutional Moment? Using Algorithmic Topic Modeling To Evaluate Bruce Ackerman’s Theory of Constitutional Change,” *Yale Law Journal* 122, no. 7 (2013): 1990–2054.

³⁴ Timothy R. Tangherlini and Peter Leonard, “Trawling in the Sea of the Great Unread: Sub-Corpus Topic Modeling and Humanities Research,” *Poetics* 41, no. 6 (2013): 725–49.

pre-understanding. This algorithmic analysis can be helpful for showing connections within or across texts which may not have been recognized by simply “reading.”

It is important to recognize that Moretti’s “distant reading,” Jockers’ “macroanalysis,” or “topic modeling” are not machine-only enterprises, with algorithms standing in the place of human close reading. A consistent critique of distant reading and topic modelling is that it promotes the false idea of the objectivity of machines performing the analysis, and thus removes human interpretation. Alan Liu, for example, expresses concern about “the meaning problem in the digital humanities,” citing topic modeling studies that assume, “A computer should be able to read texts algorithmically and discover word cohorts or clusters leading to themes without acting on an initial concept from an interpreter looking to confirm a particular theme.”³⁵ Liu is right to voice this concern, but it is important to note that much topic modeling and distant reading theory is not based on the assumption that human interpretation is removed from the process. By introducing distant reading, therefore, I am not arguing for a removal of the human’s role in reading New Testament commentaries. Moretti and many of those who follow him do not imagine that the task of reading is no longer a human task.³⁶ Jockers argues that close reading and macroanalysis, “Work in tandem and inform each other. Human interpretation of the ‘data,’ whether it be mined at the macro or micro scale, remains essential.”³⁷

³⁵ Alan Liu, “The Meaning of the Digital Humanities,” *PMLA* 128, no. 2 (2013): 414.

³⁶ “Careful reading of [Moretti’s] work reveals that this construction [distant reading] lumps together human and machine reading” (Hayles, *How We Think*, 72). For a similar response to this critique, see Goldstone and Underwood, “The Quiet Transformation,” esp. 365-66.

³⁷ Jockers, *Macroanalysis*, 26. Stephen Ramsay turns the critique of distant reading from skeptics like Liu back onto close reading. He argues that all interpretation, even close reading, is itself a recreation of a text: “The critic who endeavors to put forth a ‘reading’ puts forth not the text, but a new text in which the data has been paraphrased, elaborated, selected, truncated, and transduced. This basic property of critical methodology is evident not only in the act of ‘close reading’ but also in the more ambitious project of thematic exegesis” (*Reading Machines*, 16).

What Moretti and those who follow him call upon, therefore, is topic modeling large sets of data, to reveal patterns or developments. Understanding these patterns can then show the interpreter where close reading may be necessary. As Tangeherlini and Leonard note, “Text-mining techniques that allow for the rapid identification of ‘passages of interest’ contribute significantly to a researcher’s ability to narrow down a broader corpus into a research collection.”³⁸ This theoretical work helps to shift the notion of “reading.” In the view of Moretti, reading is less focused on distilling discrete arguments or meaning, and more focused on development. That is, both distant reading and topic modeling are diachronic approaches, as opposed to the typical synchronic understanding of reading individual works.

Topic Modelling and the New Testament Commentary

How can this diachronic approach to reading, understanding corpora of texts as systems, provide insight into the development of the genre of New Testament commentary? While this initial chapter serves to introduce the methods of distant reading and topic modeling, in the subsequent chapters I consider how to apply the ideas of Moretti to the field of the study of New Testament commentary. That is, Moretti and others provide the theoretical strategy for the approach this study proposes, but the tactics of distant reading the New Testament commentary depend upon specific instantiations of the theory. The approach to the commentary through the lens of distant reading depends upon what is distinctive about the commentary form, and this project focuses on developing “reading” algorithms that help model the topics of New Testament commentaries.

³⁸ Tangeherlini and Leonard, “Trawling,” 726.

The broad research question driving my analysis is, “What new insights can distant reading the New Testament commentary tradition offer?” This question is approached in the coming chapters through two more focused research questions relevant to the commentary genre: 1) “How would one trace the changes in the level of publication of New Testament commentaries over time?” and 2) “How would an interpreter consider how the treatment of given chapters and verses of a New Testament text has changed over time?” The analysis of the commentaries here is intended as a proof of concept, a way of previewing the value of this technique and an invitation for others to follow.

The Benefits of this Study

With this brief summary of the problem that leads to this study and the theory behind my approach in what is to follow, I close this introduction with a few notes about how I understand the importance of this study. A primary benefit of this thesis is the introduction of new reading techniques, developed outside of Biblical interpretation, into the growing field of Biblical reception history. Biblical studies in the 21st century has benefited by learning from other fields of study. For example, work drawing upon postcolonial³⁹ and reader-response theory⁴⁰ have reinvigorated the stagnating field of Biblical studies. This study is an attempt to bring in yet another field, information studies, that may help the developing field of the study of the reception history of the Bible. As noted, the close reading tradition of Biblical studies has presented a problem for analyzing a corpus as large as the set of Biblical commentaries. This

³⁹ For a number of examples, see the collection R. S. Sugirtharajah, *The Postcolonial Biblical Reader* (Malden, MA ; Oxford: Wiley-Blackwell, 2005).

⁴⁰ For a discussion of reader-response in Biblical studies, see Zoltan Schwab, “Mind the Gap: The Impact of Wolfgang Iser’s Reader-Response Criticism on Biblical Studies- A Critical Assessment,” *Literature and Theology* 17, no. 2 (2003): 170–81.

project serves to introduce Biblical scholars to the theory of distant reading and some practical ways of introducing it.

A second benefit of this study is the introduction of new sources of interpretation. These new sources come in two forms. First, distant reading is one way of avoiding the canonical problem that many distant reading theorists have identified. By reading broadly across a large set of texts, new periods and authors may become of interest. Rather than the standard commentaries that those like Childs assume are representative of a time period or distinctive from their context, mining publication data and quantifying relative treatment size may show that different commentaries, perhaps those long forgotten to history, are the true turning points of the genre. In addition to discovering new commentaries to read closely, this study incorporates data not generally found in Biblical studies. This is the metadata, of publications and authors, found in library catalogs. Few Biblical scholars are aware of the information that lies in the MARC records of their library catalogs. It is my hope that this study will show them the power of incorporating that data into their analysis of this and other genres. The investigation of metadata within textual analysis is yet another opportunity for collaborative work between librarians and textual scholars.

Third, this type of analysis will be important for showing a new way forward for digital humanities in Biblical studies. Ironically, however, this new way forward may actually be a step back. Theoreticians often speak of a second wave of digital humanities.⁴¹ Some identify an initial phase of the digital humanities that was quantitative, focused on corpus linguistics and text mining. A second wave of the digital humanities, though, is now emerging, one that is more

⁴¹ For discussion of the shifting waves in the digital humanities, see N. Katherine Hayles, *How We Think: Digital Media and Contemporary Technogenesis* (Chicago ; London: University Of Chicago Press, 2012), 24-27.

visual and aesthetic in nature. Consider the distinction made by Schnapp and Presner: “The first wave of digital humanities work was quantitative, mobilizing the search and retrieval powers of the database, automating corpus linguistics, stacking hypercards into critical arrays. The second wave is quantitative, interpretive, experiential, emotive, generative in character.”⁴² This distinction is helpful as a heuristic, but the contrast is drawn too sharply to be applicable to all areas of the humanities.

In Biblical studies, the transition from first to second wave has been reversed. That is, the most prominent digital humanities projects are those that visualize the culture of ancient Judaism and early Christianity, or digital projects that focus on re-constructing ancient cities or ancient manuscripts.⁴³ However, text mining and topic modeling have not been prominent at all in this field. This is somewhat surprising, given that the most fruitful fields for text mining have been literary studies, which often does have great impact on Biblical studies. It is my hope that the proof-of-concept function of this study will be the beginning of a conversation about how topic modeling may be a viable form of Biblical interpretation. Distant reading the history of reception and interpretation of the Bible may alert scholars to new areas of the tradition not previously studied, it may highlight shifts in scholarship not previously recognized, or it may give lie to popular perceptions of the tradition of scholarship.

It is not just that Biblical studies needs an infusion of the digital humanities. Just as Hayles notes that “the Digital Humanities should not be considered a panacea for whatever ails the humanities,” I would argue that we do not need to infuse digital projects into Biblical studies

⁴² Jeffrey Schnapp and Todd Presner, “The Digital Humanities Manifesto 2.0,” http://jeffreyschnapp.com/wp-content/uploads/2011/10/Manifesto_V2.pdf.

⁴³ For examples of the types of digital projects popular in Biblical studies, see Todd Russell Hanneken, “Digital Archaeology’s New Frontiers,” *Biblical Archaeology Review* 43.2 (2017): 28.

simply as a recipe for reviving a moribund field.⁴⁴ The digital humanities and the types of tools I am developing for this project should not be understood as simply a trend or a new set of tools. Rather, one must recognize that the birth and growth of the digital humanities is not only due to the rise of digital technologies. Rather, there are concomitant shifts in the methods of the humanities, regardless of the tools, that are playing an important role in this new form of scholarship: “The silver bullet of first-wave Digital Humanities, it seems to me, was the conjoining of a massive theoretical shift in textual studies with the practical means to implement and experiment with it.”⁴⁵ This project aims to introduce Biblical scholars to these shifts and to show them some tools that can help implement those shifts.

The history of the genre of New Testament commentary has not been written. And it may be that the difficulty of the task of understanding such a broad and long-running tradition of literature is the reason why. This study seeks to demonstrate that by learning from the field of information studies, Biblical scholars may gain new insights into this tradition of scholarship. Matthew Jockers warns, “The literary scholar of the twenty-first century can no longer be content with anecdotal evidence, with random ‘things’ gathered from a few, even ‘representative’ texts. . . . Today’s literary-historical scholar can no longer risk being *just* a close reader: the sheer quantity of available data makes the traditional practice of close reading untenable as an exhaustive or definitive method of evidence gathering.”⁴⁶ This study is an attempt to show a new way forward in the study of the history of Biblical interpretation.

⁴⁴ Hayles, *How We Think*, 23.

⁴⁵ The quotation comes from Matthew Kirschenbaum, in an interview conducted by Hayles (N. Katherine Hayles, *How We Think: Digital Media and Contemporary Technogenesis* [Chicago ; London: University Of Chicago Press, 2012], 25.

⁴⁶ Jockers, *Macroanalysis*, 9.

CHAPTER 2. READING THE BIBLIOGRAPHIC METADATA OF COMMENTARIES

“Quantitative data can tell us when Britain produced one new novel per month, or week, or day, or hour for that matter, but where the significant turning points line lie along the continuum—and why—is something that must be decided on a different basis.” –Franco Moretti¹

A driving premise of this work is that the number of New Testament commentaries makes studying the genre by close reading, the method traditionally dominant in Biblical studies, difficult. To understand the development of the genre by close reading, one is forced to choose representative samples of the genre and draw conclusions. As Matthew Jockers explores in detail, the habit of close reading often limits interpreters to “noting random things,” to identifying small parts of individual works the reader deems significant or representative of the genre as a whole.² The problem, Jockers notes, is that this approach is haphazard. Faced with a large corpus of literature, the close reader must pick and choose where to dip his or her toe, where to look for the “random things.” The result is that “something important will inevitably be missed.”³ There is an opacity and inevitable idiosyncrasy in the selection of texts to be read closely, and the result is the emergence of a “canon” of texts frequently analyzed, with many texts left unexamined due to the size of the corpus. To date, there has been little consideration of

¹ Franco Moretti, *Graphs, Maps, Trees: Abstract Models for a Literary History* (London ; New York: Verso, 2005), 9.

² Matthew L. Jockers, *Macroanalysis: Digital Methods and Literary History*, 1st Edition. (Urbana: University of Illinois Press, 2013), 8.

³ *Ibid.*, 9.

the method by which these representative samples are chosen to allow for systematic study of the genre of New Testament commentaries.

In this chapter, I introduce the mining of the bibliographic metadata of New Testament commentaries as a way of aiding in this analysis. Distant reading bibliographic metadata has the potential to reveal broad trends in the development of this genre. These trends, in turn, can offer close readers guidance in choosing where to read closely. Bibliographic metadata is a data source about the development of the commentary genre that scholars have not seriously considered. It is the function of this chapter both to introduce this new data set to interpreters.

A second function of this chapter is to demonstrate ways in which interpreters can distant read this data as a guide to understanding the development of the commentary genre. For example, one way to understand better the overall development of the genre of New Testament commentary is to understand the level of production of commentaries over time. That is, understanding the pattern of publication of commentaries may provide insight into where one should read closely in order to understand the genre. Inflection points in publication history may provide less idiosyncratic entry points for close analysis of the content of Biblical commentaries. So, the analysis in this chapter is offered as a potential answer to a key research question that is those seeking to understand the genre might ask: How has the production of New Testament commentaries changed over time? Because the size of this corpus is overwhelming and the technique of close reading is common in Biblical studies, this question has been difficult to answer. It is the function of this chapter to introduce a new method of distant reading, using bibliographic metadata, as an approach better suited to answering this question.

I begin by first discussing the methods of pulling data from the OCLC WorldCat API, including specifics about writing scripts to process the API results and some problems inherent

in relying on this data set. Then I demonstrate the power of this technique through analysis of the New Testament commentary genre by reading WorldCat metadata. I conclude the chapter with some remarks about the promise of this research method and some avenues for further work.

New Data and New Methods

I propose that the best way to analyze the bibliographic metadata of New Testament commentaries is to develop software to query machine-readable metadata, available through library catalogs, and analyze it for patterns and consistent themes. In this section, I outline the steps taken in such an approach to the present research question. In describing the tools I have developed, I also identify limitations inherent to this approach to bibliographic metadata.

1. Develop software to pull bibliographic metadata from OCLC about commentary publication, using the OCLC WorldCat API based on LCSH

The OCLC (Online Computer Library Center) is a cooperative of libraries with a mission to make information more accessible around the world.⁴ One way OCLC fulfills this mission is through its service WorldCat, a collaborative catalog of bibliographic metadata and library holdings information, constructed by the more than 16,000 members of OCLC. OCLC exposes this collaboratively-constructed bibliographic catalog by way of a web services API.⁵ This service allows software applications to request catalog records based on specific search criteria.

⁴ “About OCLC,” n.d., <https://www.oclc.org/en/about.html>. For a recent summary of OCLC’s membership and offerings, see Marshall Breeding, “Library Systems Report 2017,” *American Libraries* 48.5 (2017): 31.

⁵ For an introduction to the use of web-services APIs in libraries, see Candice Kail, “Web Service APIs and Libraries,” *Library Journal* 138.10 (2013): 120; Jason Paul Michel, *Web Service APIs and Libraries* (Chicago: Amer Library Assn Editions, 2012). For a detailed introduction to and documentation of the OCLC WorldCat Search API, see “WorldCat Search API | OCLC Developer Network,” n.d., <http://www.oclc.org/developer/develop/web-services/worldcat-search-api.en.html>.

Using defined protocols, applications can pass parameters such as publication date and subject heading to the WorldCat API and receive in response text formatted as JSON or XML.

For this demonstration, I have written software scripts to query the WorldCat catalog records based on this protocols of the WorldCat API. This mining of the WorldCat records is the first step of analysis, to be followed by scripts that parse and graph the metadata that is returned.⁶ The API is a generic web-service API, and thus such scripts could be written in any number of languages. In this study, though, the intention is to make the interface web-friendly, so that users without knowledge of software systems could make queries against the WorldCat API. With this end in mind, I have written the scripts in the server-side scripting language PHP, using AJAX to allow the user to define search criteria and asynchronously make requests.

There are limitations and assumptions inherent in this process of mining WorldCat records. Reliance upon the WorldCat search API shapes the data set I use for my analysis in significant ways. First, my analysis depends upon (and assumes) accurate catalog records in OCLC's WorldCat.⁷ While efforts are always being made to improve the quality of the WorldCat records, due to the collaborative nature of the catalog, there is no way to ensure total accuracy. For this study, I do little error correcting of the metadata I pull from WorldCat. Potential errors include duplication of records (though OCLC has de-duplicating methods in place) or inaccuracies within records (such as a recording incorrectly publication dates). There are some

⁶ All code is available at <https://github.com/boadams01>.

⁷ The accuracy of OCLC records has long been studied. For an early analysis, see Rouchton, Michael, "OCLC Serials Records: Errors, Omissions, and Dependability," *Journal of Academic Librarianship* 5.6 (1980): 316–21. For a more recent consideration of the quality of WorldCat records and initiatives to improve them, see Calhoun, Karen and Patton, Glenn, "WorldCat Quality: An OCLC Report," n.d., https://www.oclc.org/content/dam/oclc/reports/worldcatquality/214660usb_WorldCat_Quality.pdf.

normalization techniques I implement in my software to help to avoid egregious errors in catalog records. For example, in tabulating page numbers (MARC field 300 subfield a), there is a range of formats used by cataloguers. So, when parsing the Marc field 300, I run an operation to remove non-numeric characters (e.g. “pages” or “pp.”), and I reject records that have more than 2000 pages, assuming such a record is an error. Such small normalization techniques should not have significant impact on my large data set, but they are important given that there is little control over records in WorldCat.

For future research, there are some strategies that may mitigate this dependence upon collaborative records. One technique would be to draw from a single or smaller set of library catalogs. The proliferation of APIs in next generation library service platforms like ExLibris’ Alma make this a possibility for future research.⁸ For example, one could select a single, large theological library and mine its records. The benefit would be reliance upon more accurate bibliographic data, since a single institution is controlling the quality. The drawback, of course, is that such work would depend upon the individual library having records for all commentaries published, which is unlikely. Therefore, a more prudent solution might be to pull from a select set of large collections, balancing the desire for extensive records (the benefit of using OCLC WorldCat) with the desire for accurate records (the drawback of using OCLC WorldCat). Such a hybrid approach is an avenue for further research.

A second assumption about the data in this approach is that the reliance upon Library of Congress Subject Headings (LCSH) to identify works is the best way to identify commentaries. The OCLC WorldCat API allows searching by specific subject headings in catalog records.

⁸ For documentation of the API and SRU protocol for Alma, see <https://developers.exlibrisgroup.com/alma/apis>.

Fortunately, there exists a standard format for cataloguing Biblical commentaries, updated by the recent switch to Resource Description and Access (RDA). So, for the purposes of searching for Biblical commentaries, queries can safely rely on queries for subject headings in the form of “Bible. — *Book* — Commentaries,” where *Book* is replaced by the name of the Biblical book (e.g. Bible. — Romans — Commentaries). For this study, I search the OCLC API using this subject heading and making the assumption that anything with this subject heading is a Biblical commentary.

Limiting searches to this subject heading is a decision made for convenience, but it introduces bias into the data used in this analysis. The assignment of subject headings is the work of individual cataloguers, and given my reliance on the WorldCat API, the cataloguers who have created these records are distributed all over the world. Certainly different cataloguers have different criteria for considering something a commentary, and thus for applying the commentary subject heading. This challenge is exacerbated by the inherent fluidity of the boundaries of the commentary genre. That is, as noted in the opening chapter of this study, the identification of a work as a commentary is inherently a subjective decision. Consider, for example, the homilies of John Chrysostom. While typically these are referred to in titles of editions as homilies, they are pericope-by-pericope, and at times verse-by-verse, analyses of the Biblical text. In some cases, cataloguers have identified these works of Chrysostom as commentaries. For example, the homilies/commentary on the Psalms do have the LCSH subject heading “Bible.—Psalms—

Commentaries” in WorldCat.⁹ On the other hand, his homilies/commentary on Matthew do not have this subject heading in WorldCat, even though these texts are almost identical in form.¹⁰

This variation in cataloging reflects the difficulty of identifying an exact definition of what is or is not a commentary. In choosing to rely on the subject heading to identify commentaries, therefore, I introduce some imprecision into my data set. For the purposes of this analysis of publication metadata, a commentary is defined as a book catalogued with the LCSH “Bible. —*Book*—Commentaries.” Therefore, it may be most accurate to state that this analysis is focused on the trends of publications that have the LCSH subject heading “Bible. —*Book*—Commentaries.” applied to them in OCLC’s WorldCat.

2. *Parse MARC records returned from OCLC to create a database of a subset of the metadata for each commentary publication. Relevant metadata includes title, author, place of publication, date of publication, language, page numbers, book of the Bible commented upon.*

The OCLC WorldCat API returns a complete MARC record for each publication identified by the given search criteria, but this returned data has two limitations. First, by definition, the MARC record is machine-readable, and thus not in the proper format for the type of analysis I am interested in. Second, the API has limitations on the number of records returned and the number of calls that can be made in a given period of time. A second stage of analyzing the bibliographic metadata, therefore, is either to process returned records on-the-fly or store the

⁹ John Chrysostom and Robert C Hill, *St. John Chrysostom Commentary on the Psalms* (Brookline, Mass.: Holy Cross Orthodox Press, 1998). See the WorldCat record for LCSH: <http://www.worldcat.org/oclc/38966096>.

¹⁰ John Chrysostom and Philip Schaff, *Chrysostom: Homilies on the Gospel of Saint Matthew* (Peabody, Mass.: Hendrickson, 1995). See the WorldCat record for LCSH: <http://www.worldcat.org/oclc/223041543>.

returned records in a local database, and then to parse these MARC records to gather the statistics necessary for analysis of the genre.

As mentioned, the initial pulls from the WorldCat API are preformed via server-side PHP scripts. The information that is returned from WorldCat is in JSON format, and so I use JavaScript to parse the JSON, which contains the full MARC record, and gather the information I am interested in. For the purposes of analyzing the genre of New Testament commentary, there are multiple types of information that researchers may be interested in, and each requires a different approach. I identify two different approaches to the WorldCat metadata and discuss the implementation of each.

In one scenario, a researcher interested in overall consideration of the development of the genre may not be interested in the specifics of the of individual catalog records. Instead, the aggregation of records is key. For example, such a researcher may want to know the growth in commentaries on a particular book of the Bible within the decade of the 1950s. In this case, all that matters to the researcher is the publication date. Other items in the MARC record such as title and author are not important to this researcher. To run this analysis, the researcher could specify 1950 through 1959 in the search query and ask for the results to be sorted by publication date. When the results set is returned, though, he or she would need the software to progress through the results and count by year of publication.

In a second scenario, more details of returned records are important. Consider, for example, the scenario in which a researcher is interested in the average number of pages for commentaries of a given book over time. The query might be the same as the scenario above, but in this case the specific number of pages (MARC field 300 subfield a) would be important. The

user would need software to process through the records and keep a running tab of the average number of pages for the result set.

Results from these two use scenarios are detailed below. In each case, though, the approach of the software may be slightly different. For demonstration purposes, I have developed three ways of handling the results of WorldCat API calls. The first is a manual process. I have written scripts that count the number of publications based on a set of criteria (for example, the year of publication). The user can export the raw data returned from WorldCat. The user then has the option to import that raw data into a separate analysis engine. In this demonstration, I have imported the results of the queries into the visualization engine Tableau to create visualizations.

A second approach, more automated for the user, is to chart the data as it is returned from the WorldCat API. In this scenario, the software passes the the data to the Google Charts API, which will return visualizations automatically.¹¹ While these visualizations are less sophisticated than those produced through Tableau, they require nothing of the user.

A third alternative is to store the data returned from WorldCat locally in a SQL database. The reason one might prefer this option is to overcome two specific usage limitations of the WorldCat API.¹² First, OCLC limits the number of records a single query can return to 100. For the previous two methods of handling the data, I overcome this limitation by making repeated calls and storing the results locally in memory before processing. An alternative, though, is to separate the call and the processing, and thus to store the results in a more persistent form, the database. A second limitation imposed by OCLC is the limit on the number of calls (50,000) that

¹¹ For full documentation of the Google Charts API, see <https://developers.google.com/chart/>.

¹² For documentation of the limitations on the WorldCat API, see <https://www.oclc.org/developer/develop/web-services/worldcat-search-api/faqs.en.html>.

can be made with one API key in a single 24-hour rolling period. As the dataset of commentaries to be analyzed grows, the user runs the danger of running over this limit in the first two methods. Therefore, as an alternative the software can store results locally as they are returned from OCLC. The danger of this approach is that any analysis is then running on stale data, rather than pulling from WorldCat at the time of analysis. The assumption of this process, though, is that WorldCat catalog records are stable enough to reasonably delay analysis. For future analysis of the dataset stored in the database, it would be important to refresh the locally-stored records before running the analysis.

All three of these methods are viable approaches to processing the OCLC WorldCat metadata, and each has its benefit. The first, dumping the raw data, allows the sophisticated user to import the data into his or her visualization engine of choice.

Initial Results of a Preliminary Analysis of OCLC WorldCat Metadata on New Testament Commentaries

This enumeration of the method of this study of bibliographic metadata and the assumptions inherent is offered as a guide to future study, introducing bibliographic metadata as a source of analysis and showing the ways of implementing distant reading with software. I now turn now to offer some initial results from distant reading the bibliographic metadata of New Testament commentaries. I emphasize that this study is a proposal of a method and an invitation for further distant reading studies of bibliographic metadata. Therefore, these examples should be taken as just that, examples of the power of this type of analysis. Tracking the metadata of commentaries demonstrates very clear trends in the development of this genre. In what follows, I outline some important tendencies in the development of this genre since 1901. Each of these

trends may help identify locations where close reading may bear fruit. In the final section of this paper, I offer some invitations for further study.

The Rise of the Modern Commentary “Industry”

OCLC bibliographic metadata clearly shows considerable growth in the number of New Testament commentaries published during the 20th century. Most New Testament scholars would suggest that this growth is evident by simply perusing the catalogs of major publishers or touring the book displays at Biblical studies conferences. David Clines refers to the “plethora of commentaries, good, bad, and indifferent, that publishers insist on setting before us,” but he does so without citing any specific evidence of this plethora.¹³ Publication data available through OCLC WorldCat bibliographic metadata, though, offers firm support for this anecdotal evidence.

The OCLC metadata demonstrates a sharp increase in the number of commentaries published in the last half of the 20th century. When looking at the aggregate number of commentaries published for all 27 books of the New Testament, one sees a marked rise that begins after World War II. From 1901 through 1949, there were an average of 120.49 commentaries published each year on all books of the New Testament. From 1950 through 2016, though, that average rose to 424.43.



Figure 1: Publication Levels of Commentaries on all Books of the New Testament

¹³ David J. A. Clines, “Writing a Job Commentary,” in *The Genre of Biblical Commentary* (Eugene, OR: Wipf and Stock, 2015), 38.

This Post-WW2 rise is consistent across the New Testament canon. Consider, for example, how a graph of commentaries written on the four New Testament gospels mimics the shape of the graph for all books of the New Testament.

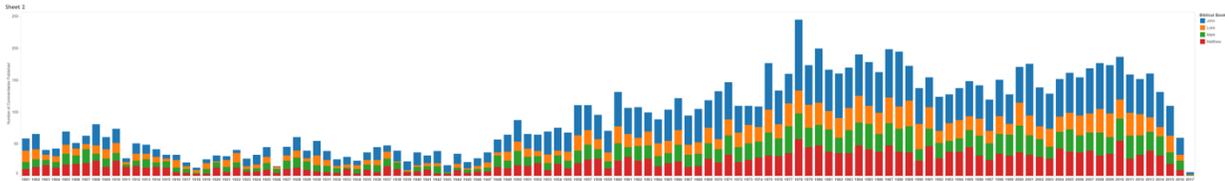


Figure 2: Publication Levels of Commentary on the Four Canonical Gospels (1901-2017)

There are, however, some notable exceptions across the canon. For example, the rise in commentaries on Matthew, Acts, and Hebrews is clearly demarcated in the late 1940s, but the growth number of commentaries on 1 Corinthians (the blue line below) is less pronounced in the 1940s. Instead, the number of 1 Corinthians commentaries stays relatively low until the 1960s.

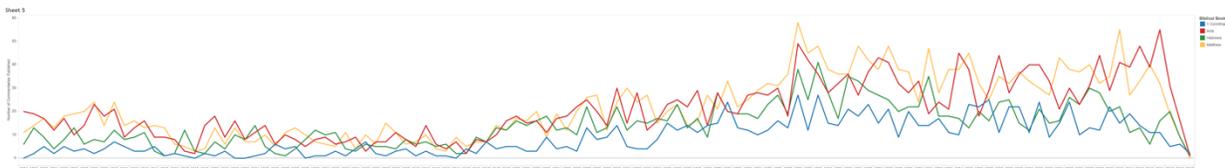


Figure 3: Publication Levels of Commentaries on Matthew, Acts, 1 Corinthians, Hebrew (1901-2017)

As this chart demonstrates, there is a significant spike in commentaries published in the year 1978 for most books of the New Testament. In 1978 there were 775 commentaries published, a huge number in comparison with 457 in 1977 and 548 in 1979.

The data outlined here suggest that scholars would be well-served to look closely at the New Testament commentaries of the 1940s and the 1950s. Are there substantive differences between the commentaries produced in these decades and those that preceded them? Here the

distant reading of bibliographic metadata serves as a guide to suggest places where close reading may be productive.

There has been general stability and yet some changes in the relative popularity of books of the New Testament

Publication metadata reveals the relatively consistent popularity of certain books of the New Testament for commentators. From 1901 through 2017, the most commented upon book of the New Testament has been the Gospel of John (4,292 publications), and the least commented upon book has been 3 John (63 publications). These results will not be surprising to someone who understands the reception history of the New Testament. What may be surprising to some New Testament scholars, though, is that the book of Revelation has been a close second (3,916 publications). There are been several years, in fact, when more commentaries on Revelation were published than commentaries on John.¹⁴ After 1951, though, John has been fairly consistently the most popular book for commentaries, though Revelation has been close in popularity.

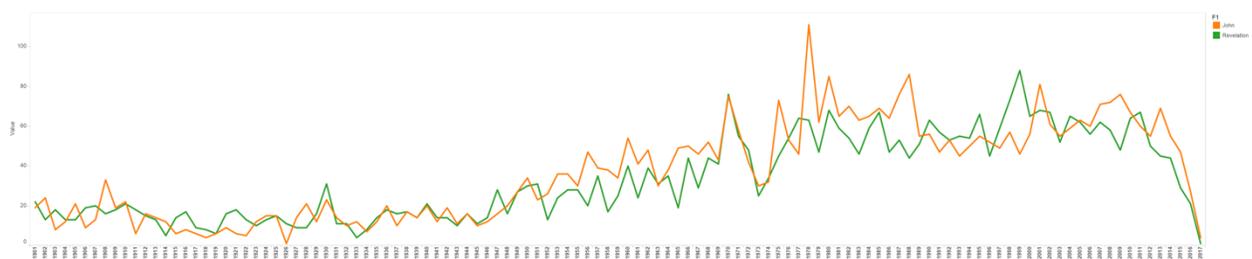


Figure 4: Publication Levels of Commentaries on John and Revelation (1901-2017)

¹⁴ See, for example, 1930 (31 commentaries on Revelation v. 23 on John), 1947 (28 v. 16), and 1999 (88 v. 46).

In terms of the four New Testament gospels, John has always been the “favorite” of the guild of New Testament scholarship. After John, Matthew has been the second most popular (2,566 publications), followed by Mark (2,188) and Luke (2,111). This divide between John and the synoptic gospels grew throughout the 20th century, and the divide continues to grow.

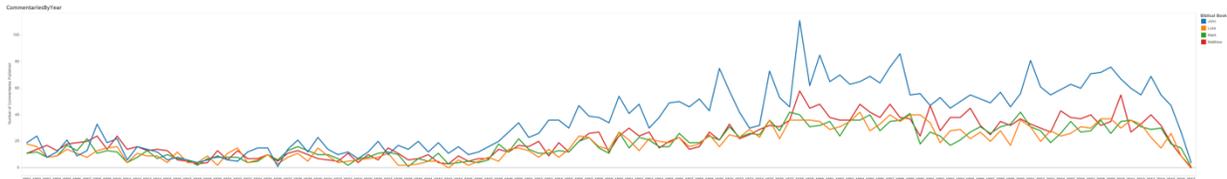


Figure 5: Publication Levels of Commentaries on Each of the Four Canonical Gospels (1901-2017)

If one looks at earlier data, though, from the 19th century, the clear shift toward Mark becomes evident, coinciding with the 19th-century development of Markan priority, the argument that rather than being derivative of Matthew, Mark was the first gospel written, serving as a source for Matthew and Luke.

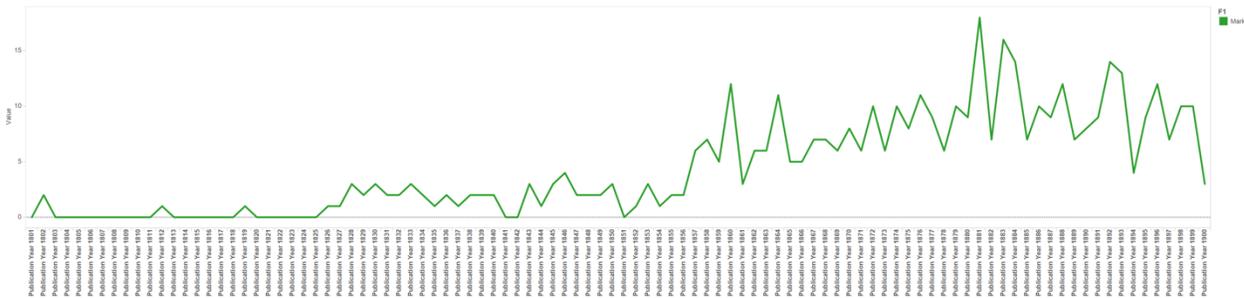


Figure 6: Publication Levels of Commentaries on Mark (1801-1900)

Likewise, in the study of the thirteen letters of Paul, the data show that Romans has always dominated in terms of the number of commentaries published. This result should not be surprising to most New Testament scholars. What may be surprising, though, is the relative importance of Galatians and Ephesians, more so than Philippians.

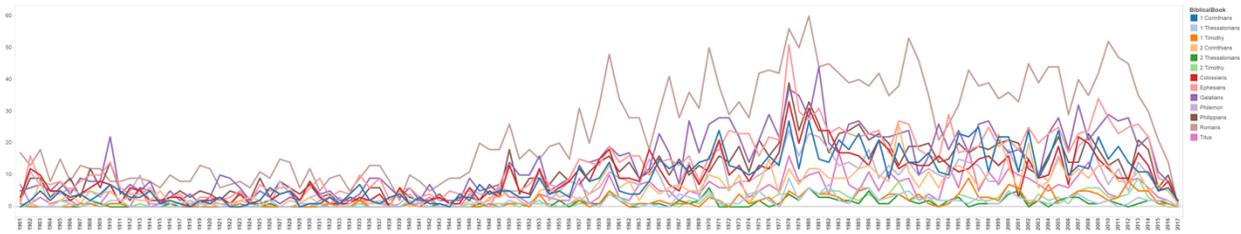


Figure 7: Publication Levels of Commentaries on Each of the Thirteen Pauline Letters (1901-2017)

Throughout the 20th century, scholars' preference for a few Pauline epistles (Romans, Galatians, Ephesians, Colossians) became more pronounced, while the number of commentaries on other letters remained relatively flat. Consider, for example, commentaries on Galatians versus commentaries on Philippians, from 1941 through 1970, the period of intense growth in New Testament commentaries.

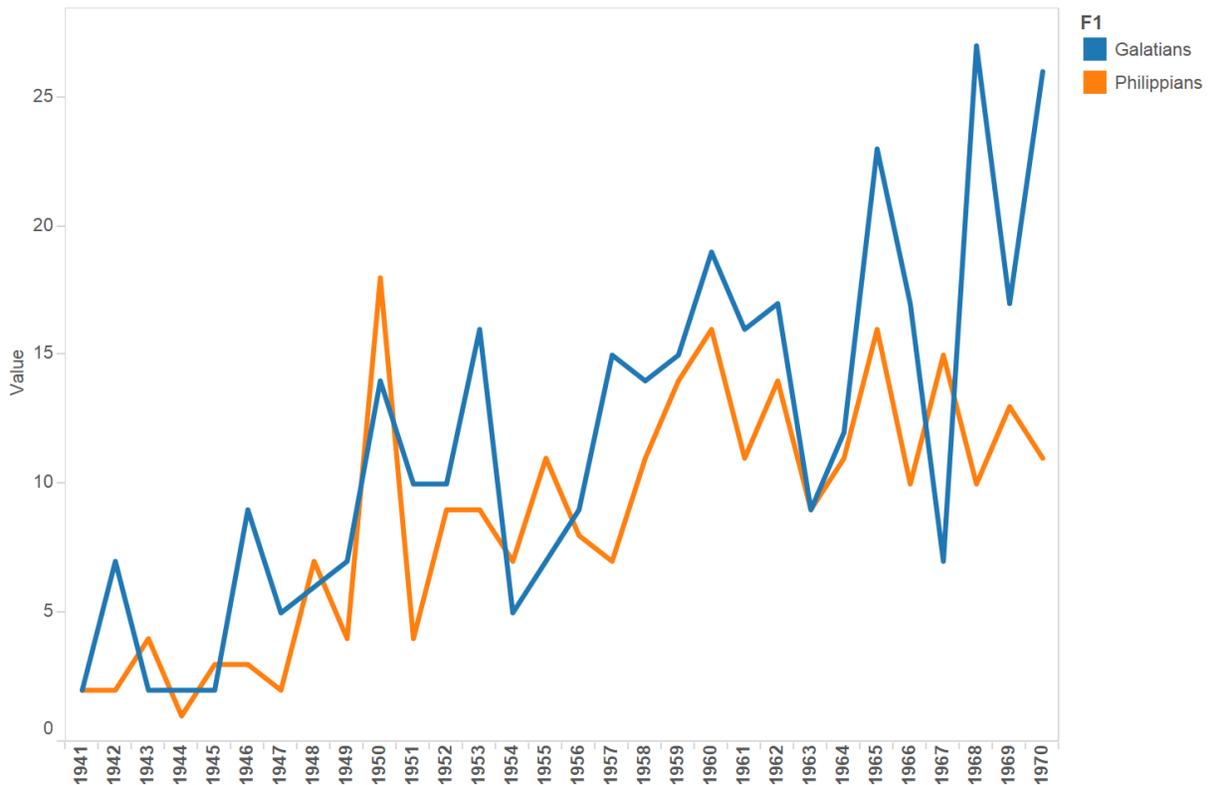


Figure 8: Publication Levels of Commentaries on Galatians and Philippians (1941-1970)

The data also reveal some interesting trends among the so-called disputed Pauline epistles.¹⁵ Ephesians and Colossians have been the subject of a growing number of commentaries, while the number of commentaries published on the other 4 disputed letters has remained relatively flat compared with the general growth in New Testament commentaries, even since the 1940s when there was marked growth in commentaries on the rest of the New Testament canon.

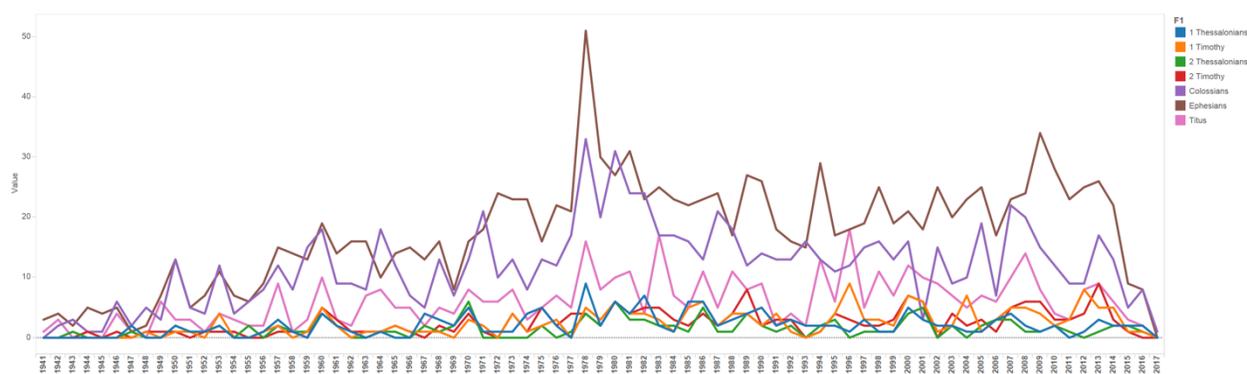


Figure 9: Publication Levels of Commentaries on the Disputed Pauline Letters (1941-2017)

These few examples should make clear of the types of patterns that mining the OCLC metadata can reveal. These numbers cut against the grain of some assumptions of New Testament scholarship and may suggest productive avenues of further study.

The Growth in Publication of New Testament Commentaries is Consistent with the Growth in New Testament and Old Testament Studies, but it Differs from Other Fields

That the post-WW2 decades were a period of growth in humanities scholarship is not entirely surprising. OCLC metadata for any number of LCSH shows a general increase in the number of publications in a range of fields. It is important to note this trend, and to recognize the

¹⁵ The disputed Paulines include Ephesians, Colossians, 2 Thessalonians, 1 Timothy, 2 Timothy, Titus.

danger of drawing absolute conclusions from the OCLC metadata. Perhaps it is the case that there is a general growth in the number of catalog records, which would account for much of the documented rise in the number of commentaries published.

To understand how much of the growth in this genre documented here is the result of a general growth in WorldCat records, it is important to study the rise in publications catalogued with other LCSH. Such analysis shows, in fact, that the growth of publications in the field of Biblical Studies does not mimic growth in publications in literary studies of other corpora. To demonstrate this point, I have considered subject headings of other religious publications (Augustine) and other popular literary authors (Melville) and literary texts (Shakespeare’s “Hamlet”). The marked growth in New Testament and Old Testament studies far outpaces the 20th-century growth of the study of Shakespeare, Augustine, and Melville’s *Moby Dick*.

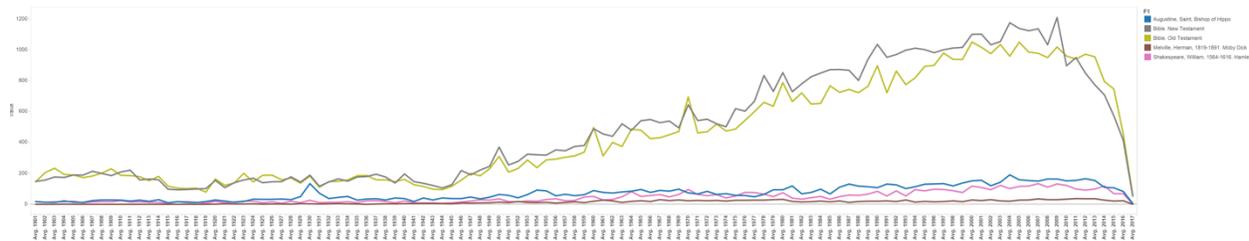


Figure 10: Growth of OT and NT Studies Compared with Other LCSH (1901-2017)

Consider, for example, the comparison of the rise of the New Testament commentary to the rise of publications on Shakespeare’s “Hamlet.” The number of publications with the subject heading “Shakespeare, William, 1564-1616. Hamlet” was 16 in 1901 and 99 in 2011. That is growth of 618.75% in publications. In the similar period, the number of publications with the subject heading “Bible. New Testament” grew from 148 in 1901 to 1103 in 2011, growth of 745.27%. This growth, though, does not occur at the same rate or at the same time period. Consider the average number of publications by decade for several humanities-based subject headings:

LCSH	1901-1910	1911-1920	1921-1930	1931-1940	1941-1950	1951-1960	1961-1970	1971-1980	1981-1990	1991-2000	2001-2010
Bible. New Testament	184.5	134.6	133.0	164.1	188.7	344.4	520.1	643.3	858.4	1005.5	1091.5
Shakespeare, William, 1564-1616. Hamlet	16.3	13.5	12.9	15.8	17.2	28.9	56.9	62.2	51.8	85.3	116.2
Augustine, Saint, Bishop of Hippo	19.7	19.3	26.3	40.8	39.4	66.9	83.5	70.0	101.1	127.8	155.3
Melville, Herman, 1819-1891. Moby Dick	0.1	0.1	2.0	2.6	7.1	13.2	22.6	26.2	18.0	20.2	27.9

Figure 11: Number of Publications by Decade for Given Subject Headings (1901-2010)

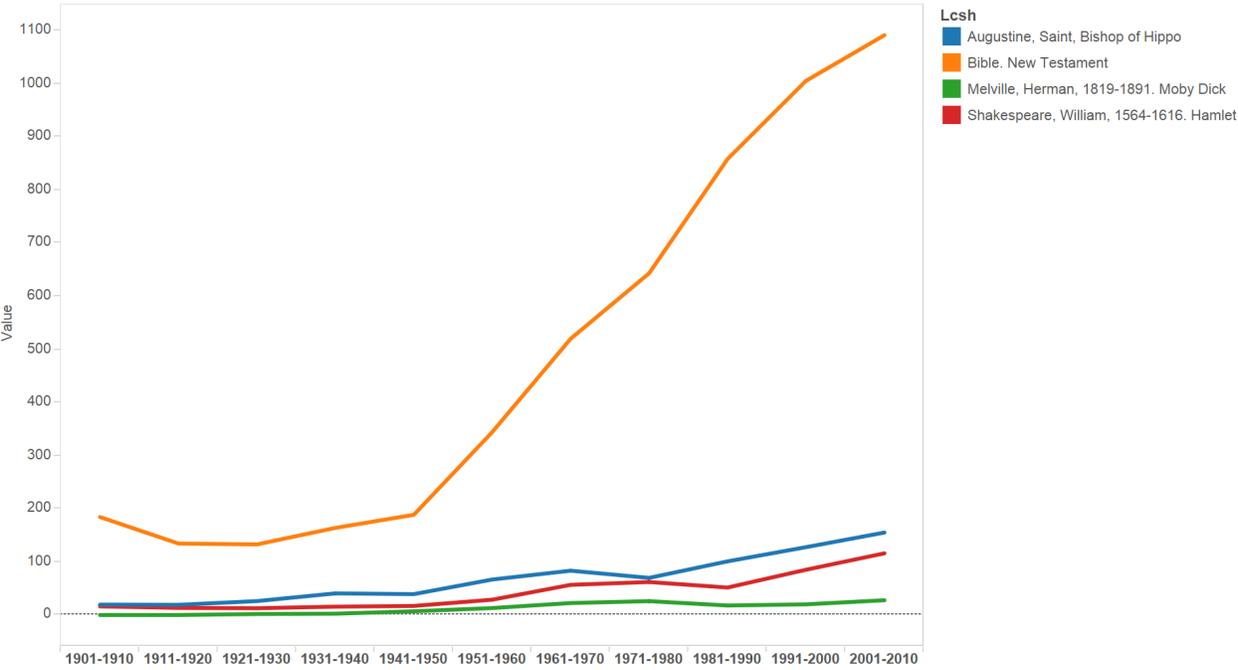


Figure 12: Publication Levels for Works with Given Subject Headings (1901-2010)

Therefore, it is evident that the growth in publications about the New Testament (LCSH “Bible. New Testament”) far outpaces growth in publications in other fields of humanities study.

This would suggest that the increase in commentaries published post-WW2, noted above, is not primarily attributable to the growth of publications in general or the growth in catalog records available in WorldCat.

The Growth in the Size of Commentaries

The demonstration of reading bibliographic metadata to this point has focused on tracing the number of commentaries published over time. However, the WorldCat API returns a full MARC record, and thus one can use the API to follow other trends in publication history. In this section I demonstrate an alternative mode of analysis, looking not at counting commentaries by publication date, but rather performing analysis on metadata related to the content of the commentary itself. To perform more complex operations on the bibliographic metadata, in this case averaging page numbers of individual publications, it is necessary to download OCLC records and store them locally in a database. This section, therefore, demonstrates not only an alternative use of metadata, but a different software approach to the process of mining.

One particular part of the MARC record ripe for analysis is the 300 field of the MARC record, the physical description of a volume. Included in the physical description is the number of pages of a work. While the page numbers of a work are not determinative of its content, one can draw certain conclusions about the content of a commentary based on how long it is. As noted in the introduction to this study, the size of recently-published commentaries seems to have increased tremendously. Reading the OCLC metadata of commentary publications over time, combining the 264 field (publication date) with the 300 field (physical description), allows readers to test whether this is actually true. OCLC metadata substantiates this perception and

allows interpreters to identify inflection points in the trend, which may suggest something about changes in approach to a commentary.

The expansion of analysis to fields other than the 264 field (publication date), though, brings new challenges. First, the field is more open ended in terms of what is entered into it. For example, subfield a of the 300 field may include the page designation of front matter, often in Roman numerals (but not always) or an identifier of “pages” or “pp.”, or the subfield a may be left blank completely. Therefore, when mining and processing this data, more normalization techniques must be put into place. In the examples below, the software excludes non-numeric characters, removes items for which the subfield a is blank (or includes all nonnumeric characters), and then excludes records with illogical results for the number of pages (the software will overlook items with less than 10 or more than 2000 pages).

Second, because the analysis is more complex drawing upon multiple MARC fields (in this case 264 and 300), it is more difficult to handle on the fly, as the records are pulled. For demonstration purposes, therefore, I have downloaded the OCLC records into a local database and then processed them. This is not completely necessary (analysis of the 300 field could happen as the records are pulled), but this demonstrates another possibility for analysis that is more complex.

With these design changes in place, though, the bibliographic metadata does show some significant changes in the size of New Testament commentaries produced over the last few decades. Consider, for example, the size of commentaries on the book of Revelation from 1950 through 2016.

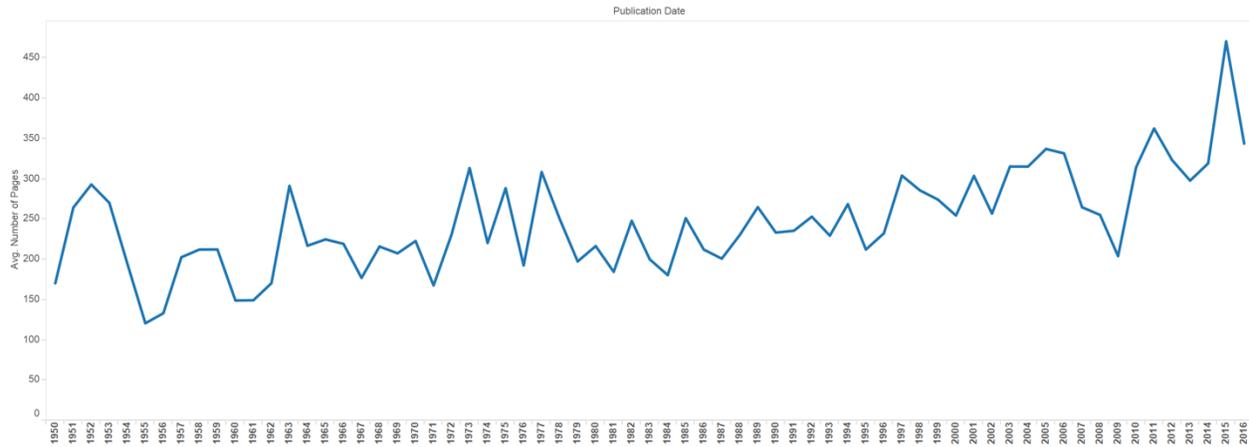


Figure 13: Average Number of Pages for Commentaries on Revelation Published within a Given Year (1950-2016)

Since 1950, the average number of pages in commentaries on Revelation published during a given year has risen by over 100 pages, averaging 214.89 pages in the 1950s and averaging 341.15 pages from 2011 through 2016.

This trend of growth, found in Revelation, is less pronounced, though, in commentaries on other books of the New Testament. Consider, for example, commentaries on the gospel of Mark, which shows a slight upward trend, but is more flat than the chart of Revelation page numbers.

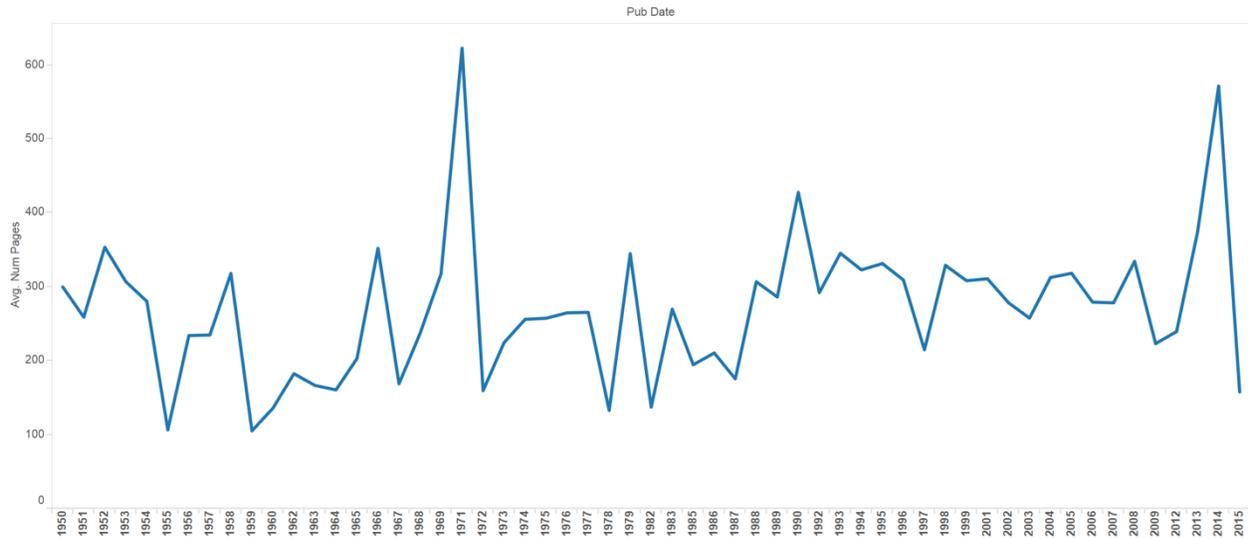


Figure 14: Average Number of Pages for Commentaries on Mark Published within a Given Year (1950-2015)

The Mark data shows an interesting spike in 1969, which is not the result of a single commentary, but of several commentaries of long length that year. Moving back in time, we see that before 1950, the variability in the number of pages for Mark commentaries is higher year over year.

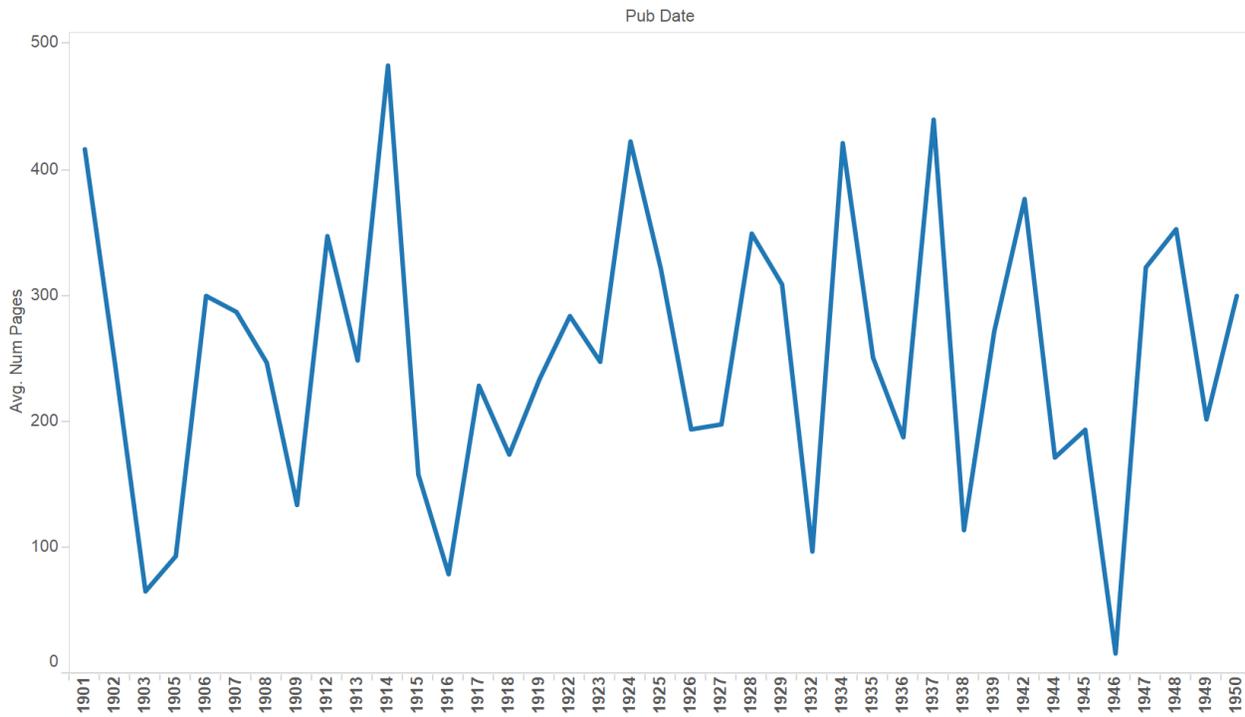


Figure 15: Average Number of Pages for Commentaries on Mark Published within a Given Year (1901-1950)

Taking a broad look at the size of commentaries over time can also reveal works that stand out, which may have not received sufficient treatment. Consider, for example, the average number of pages of commentaries produced each year on the book of Acts:

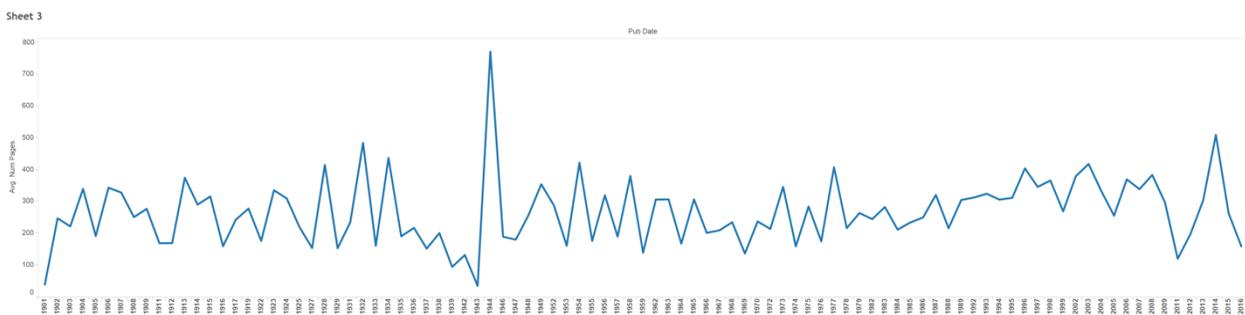


Figure 16: Commentaries on Acts Published by Year (1901-2015)

The trend of general growth is evident (an average of 257.09 in 1901-1910; an average of 353.26 in 2001-2010), but what stands out is the anomalies, such as the publication year 1944. Looking

closely at the data, we find a single publication, the 1,134 page commentary of Richard Charles Henry Lenski, accounting for this anomaly.¹⁶ Print copies of this commentary, though, are only held by 2 libraries, suggesting the work is not often read.¹⁷ The analysis here suggests that this commentary may have represented an approach uncommon in the 1940s, and therefore perhaps it warrants more attention than holdings data would suggest it is receiving.

These brief demonstrations of the change in the size of commentary publications demonstrate the potential power of analyzing the combination of MARC fields to understand the trend of New Testament commentaries. Scholars' general perception that the size of commentaries has grown can now be substantiated by actual bibliographic metadata. This substantiation allows scholars to investigate when this change occurred, whether there are particular interpreters who prompted this change, and thereby reassess which are the representative samples of the genre that merit close reading.

The Use of the OCLC WorldCat Search API and the Study of the New Testament Commentary

The analysis here is offered as an introduction of OCLC metadata as a data source for understanding the development of the genre and a demonstration of how to distant read this bibliographic metadata. The focus here is on introducing the approach to OCLC metadata, rather than drawing substantial conclusions about the development of the genre. Still, this quick analysis shows significant changes in the commentary genre in the 20th century, and it suggests opportunities for scholars to look closely at a few years or few individual commentaries as points of change. The significance of this analysis is that the determination of those years and

¹⁶ R.C.H. Lenski, *Interpretation of the Acts of the Apostles* (Columbus, O.: Wartburg Press, 1944).

¹⁷ For holdings information, see the OCLC record: <http://www.worldcat.org/oclc/878554973>.

commentaries is based on the data, rather than the idiosyncratic choice of the scholar. The analysis here, though, is offered primarily as an invitation for others to explore this as a method of understanding the development a genre which is so difficult to understand. In conclusion, I would like to introduce a few possibilities for how this research could be extended further.

First, in this example I used metadata that is available through the OCLC WorldCat API itself. The power of APIs, though, is the ability to combine data sources and to read OCLC metadata in combination with other more specific sources. In this case, the WorldCat API is limited. Consider, for example, the mining of author data. One might imagine that one could learn much about the development of the genre by asking questions about who is writing these commentaries. Are commentators men or women? Are they protestants or Catholics? How old is the average commentator? These are all questions that the metadata could conceivably answer, but the OCLC metadata is not situated well to do so. Though the WorldCat API can pull author information from the 100 field of the MARC record, this field is simply encoded as a text field, often including an author's name, with a single spelling without alternatives, and generally a birth and death date. Such limited information is not helpful in answering the questions posed above. Imagine, however, taking the author information for New Testament commentaries and combining that information with a resource like the Virtual International Authority File (VIAF) or the Library of Congress Name Authority File (LCNAF).¹⁸ This combination would allow one to run analysis on demographics of the authors of commentaries. What is the breakdown of gender in who is writing the commentaries? Has that balance shifted over time? Are younger

¹⁸ VIAF became a service of OCLC in 2012. For a discussion of the relationship, see Michael Kelley and Meredith Schwartz, "Viaf Service Transitions to Oclc," *Library Journal* 137.8 (2012): 16. For a discussion of the potential for the use of VIAF in LIS, see Thomas B. Hickey and Jenny A. Toves, "Managing Ambiguity In VIAF," *D-Lib Magazine* 20.7/8 (2014): 18–24.

scholars writing more commentaries? Combining the OCLC metadata with other sources would allow for more in-depth analysis. One might imagine that even more specialized databases of Biblical scholars exist, and combining the MARC 100 field with these records, through authority files, could enrich the analysis of the genre.

Unfortunately, at this point such combination would be imprecise and require manual verification. Comparing the text of the 100 MARC field to a VIAF entry is not an exact science. For example, one author of a recent commentary on Revelation is David E. Aune, who has a VIAF entry that is accessed via the VIAF API:

```
http://www.viaf.org/processed/search/processed?query=local.personalName+all+%22david+aune%22%20&recordSchema=info:srw/schema/1/briefmarcxml-v1.1&maximumRecords=5&httpAccept=text/xml
```

The resulting XML has authoritative records for the author's gender (field 375) and birthdate (field 046). This query, though, depends upon a precise search term (in this case "David Aune" as the personal name), pulled from the WorldCat API. A far more effective system would be one in which the entry in the OCLC record for the commentary's author was not the string "David E. Aune," but rather a link to the VIAF entry or some other authority file.¹⁹

The cataloguing world, though, is moving in a direction that will allow for this integration of data sources to happen with precision and automatically. This integration is the promise of linked data, the creation of connected data sets on the web. So, for example, instead of using a text field of the author's name to identify the source of a work, in a linked data model, that source would be a uniform resource indicator, not the text of an author's name, but rather a

¹⁹ The VIAF record includes URIs for other bibliographic sources. For example, the Aune record links to the Katalog der Deutschen Nationalbibliothek (<http://d-nb.info/gnd/182790908>).

link to an authority file that contains information about the author.²⁰ Using the APIs to both WorldCat and the VIAF, connected by linked data, analysis of the development of the New Testament commentary could be far more complex and more revealing of patterns and trends, helping scholars to identify where best to apply close reading.

This move toward linked data is an example of where the ongoing work of librarians to improve metadata will allow subject-matter experts to better understand their subjects. It also demonstrates, though, how this collaborative effort can go in the other direction. That is, cataloguers may need to consider how they are organizing metadata in light of changing research needs. As modes of reading like the one demonstrated here become more popular, metadata may need to be better-organized to be truly machine-readable. Librarians and researchers must be aware of the work each is undertaking, so that they may mutually support new work.

Second, this analysis shows the need for studying the growth of OCLC WorldCat records themselves. That is, I have shown that the growth in commentary publications over time is consistent with a general growth in the number of records in OCLC for many subject headings. As noted, the growth in Biblical studies is far more pronounced, a fact consistent with anecdotal evidence of the expansion of Biblical studies. Still, it cannot be denied that some of the development of the genre is the result of more records from more recent publications being available in WorldCat. It would be helpful for this type of analysis to understand how WorldCat itself has grown over time. This introductory study has shown that there is a need for a systematic control set to account for how much WorldCat itself has grown.

²⁰ OCLC is moving quickly toward linked data. For a discussion of the future, see Karen Coombs, “WorldCat Discovery API and Linked Data | OCLC Developer Network,” n.d., <https://www.oclc.org/developer/news/2014/worldcat-discovery-api-and-linked-data.en.html>.

Third this analysis shows there is a need to explore better what constitutes a commentary. In this study, I have relied on the Library of Congress' Subject Headings to identify what a commentary is. However, a more systematic approach would draw upon a list of items that the guild itself has recognized to be a commentary. Though this list would be large, it is not so large that it could not be formulated comprehensively. WorldCat may be the best metadata source for commentaries, but relying on subject headings is problematic, as noted above in the identification of cataloguers' idiosyncratic preferences. To study fully this genre, Biblical scholars need to agree on what constitutes a commentary and then identify the set of commentaries that have been published over time. The OCLC accession number is an ideal way of keeping track of these publications, as that would allow researchers to centralize the metadata on these publications. To improve the analysis here, though, it would be helpful to know what constitutes a commentary in the history of Biblical interpretation.

I offer one final concluding suggestion for further study as a segue to the next chapter of this study. Analysis of bibliographic metadata suggests that there are inflection points in the development of this genre. That is, there are texts or clusters of texts that mark significant changes in the genre. As noted above, for example, OCLC metadata shows that Lenski's 1944 Acts commentary stands apart from its context in terms of its length. Why was this one commentary so different? As recommended above, this commentary might be a good place to return to the traditional form of analysis in Biblical studies: close reading. However, this outlier might also be a good candidate for digitization, so that the distance analysis introduced in the next chapter could be applied to it. As I will explore in the next chapter, it is surprising how little of the commentary tradition has been digitized. This lack of digitized text is a major hindrance to distance reading techniques. It is not that commentaries have not been digitized. Rather, it is that

what one finds digitized is haphazard; there is no clear pattern as to what has been digitized, even for public domain materials. The analysis introduced here may serve as a new guide for where digitization efforts should be applied. The analysis has shown rather clearly, for example, strong growth in the commentary tradition in the 1940s. It would stand to reason, therefore, that the 1940s should be a primary focus of digitization efforts. I am hopeful that this method of reading bibliographic metadata could be used as an initial guide to where digitization efforts should be more intense. Perhaps combining the WorldCat metadata with metadata available from other digital books sources, such as the Internet Archive, might be a fruitful way of suggesting where digitization should take place.²¹

It is clear, therefore, that there is much to be learned by reading the bibliographic metadata of commentaries. It has been my intention to demonstrate how this might be done, in hopes that others will see this alternative form of “reading” as a productive mode of studying the history of Biblical reception. In the next chapter, I turn to a more formal approach of distant reading the genre.

²¹ The Internet Archive does have an API that can be used to pull metadata for items in their collection. For documentation, see <https://archive.org/help/json.php>.

CHAPTER 3. READING COMMENTARIES AS STRUCTURED TEXTUAL DATA

“Some books are read for immersion—the familiar and comfortable image of a reader under a tree or in some other naturalistic setting is a reflection of the ‘silence and slow time’ we typically associate with such reading, that is the deep, meditative pleasure of becoming ‘lost’ in a book. But not all, indeed not most, books are destined for this kind of reading.” –Matthew G. Kirschenbaum¹

In the first chapter of this study, I identified a commentary as a distinct literary genre based on form. Despite their vast differences in hermeneutical approaches and theological assumptions, there is a connection between between the New Testament work of Hippolytus of Rome (early 3rd C CE) and Luke Timothy Johnson (early 21st C CE), and that connection is the literary mode of offering comments on individual sections of the New Testament.² The genre of the New Testament commentary is remarkable in the stability of its form.

It is the function of this chapter to show that this stability of form provides an opportunity for scholars interested in analyzing the development of the commentary genre. The commentary is a specialized form of literature wherein the form is directly connected to the content. That is, a

¹ Matthew G. Kirschenbaum, “The Remaking of Reading: Data Mining and the Digital Humanities,” *Conference Proceedings; National Science Foundation Symposium on Next Generation of Data Mining and Cyber-Enabled Discovery for Innovation, Baltimore, October 10-12, 2007* (n.d.), <https://www.csee.umbc.edu/~hillol/NGDM07/abstracts/talks/MKirschenbaum.pdf>.

² It should be noted that verse numbers were a later addition to the Biblical text. However, editors have sense gone back and divided ancient commentaries into the modern Biblical verse system. For a discussion of the introduction of verse numbers into the Biblical text, see Nicholas J. Zola, “Why Are There Verses Missing from My Bible? The Emergence of Verse Numbers in the New Testament,” *Restoration Quarterly* 54.4 (2012): 241–53.

commentary is already textual data, structured by the canonical arrangement of chapter and verse. Given this inherent structure to the text, readers can discern meaning or focus of the commentary based on the form of the work. A second distant approach, described in this chapter, invites interpreters of the genre to trace shifts in emphasis across the commentary tradition, tracking which verses and chapters of New Testament texts attract more or less attention from commentators as the genre develops. Using this stability of form, readers can compare commentaries that diverge wildly in their time of publication or method of interpretation and therefore better understand the developing conversation around particular New Testament texts.

To understand the need for this method, consider the challenge of comparing two commentaries on 1 Corinthians, one published by John Calvin in 1546 and one published by Gordon Fee in 2014.³ Though both are commentaries on the same Biblical text, close comparison between these two would prove difficult, given the size of each (Calvin's commentary is 158,260 words; Fee's is 217,773 words) and the fact that these two interpreters are products of very different contexts of scholarship, and thus their approaches to Paul's letter to the Corinthians are very different. Where should one begin such a comparison? This challenge is compounded if one is not only interested in reading these two alongside one another, but understanding how the genre developed in the four and half centuries between their publication, across the thousands of commentaries published on this same Biblical text.

By taking advantage of the consistent verse-by-verse form of the two commentaries, though, readers can "step back" from the details of these massive tomes and see quickly where productive points of comparison might lie within these treatments of the letter. If one breaks

³ Jean Calvin et al., *The first epistle of Paul the apostle to the Corinthians* (Grand Rapids, Mich.; Carlisle, Cumbria, U.K.: Eerdmans ; Paternoster, 1996); Gordon D Fee, *The First Epistle to the Corinthians* (Grand Rapids (Mich.); Cambridge: W.B. Eerdmans, 2014).

down the relative size of treatment of each verse of the letter, interesting differences emerge that may provide the entry point for comparison. As a control, consider that each of the 16 chapters of 1 Corinthians receives roughly the same relative treatment in these two works, based on the number of words dedicated to the chapter as a percentage of all the words in the commentary. Both Calvin and Fee give relatively the same amount of attention to each chapter of Paul’s letter. This similarity is shown clearly in Figure 1.

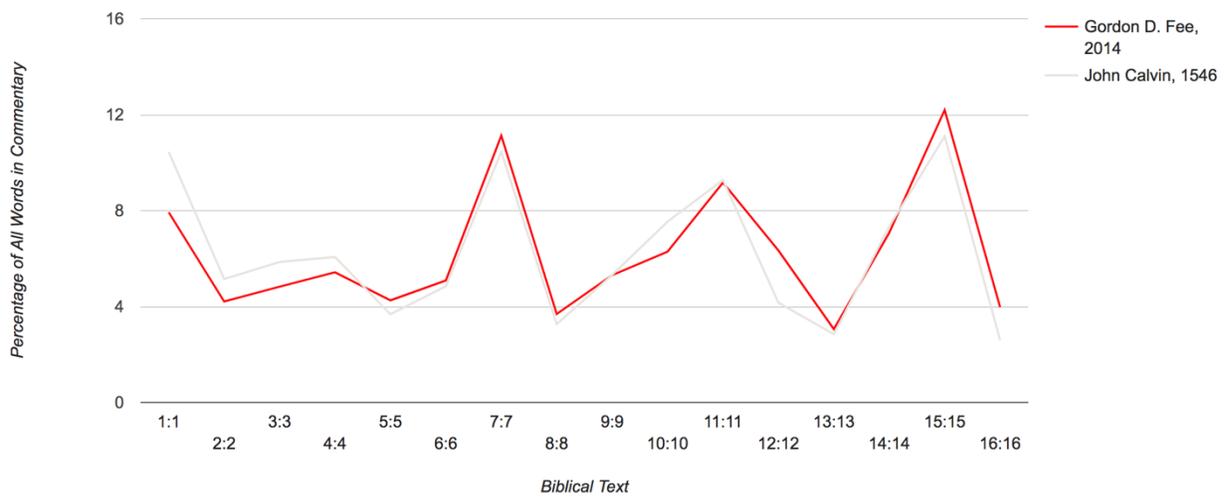


Figure 17: Comparison of the Relative Treatment of Each Chapter in 1 Corinthians, Fee and Calvin

However, if a reader looks more closely at the relative size of the treatment of the individual verses of the letter, readers notice pronounced differences between the focus in these two commentaries; it becomes evident where the points of emphasis in each treatment may lie.

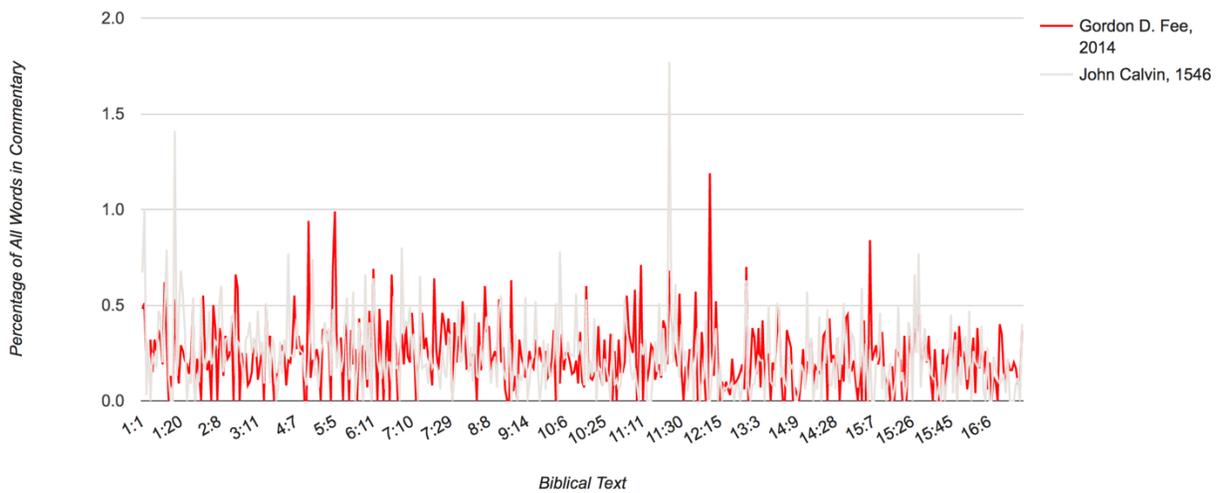


Figure 18: Comparison of the Relative Treatment of Each Verse in 1 Corinthians, Fee and Calvin

Figure 2 shows a clear spike in Calvin's attention at the end of the 11th chapter of 1 Corinthians. Mapping the relative treatment of the 34 individual verses in 1 Corinthians 11, Figure 3 shows that Calvin gives far more attention to v. 24 (1.77% of the total words in his commentary) than does Fee (0.68% of the total words in his commentary).

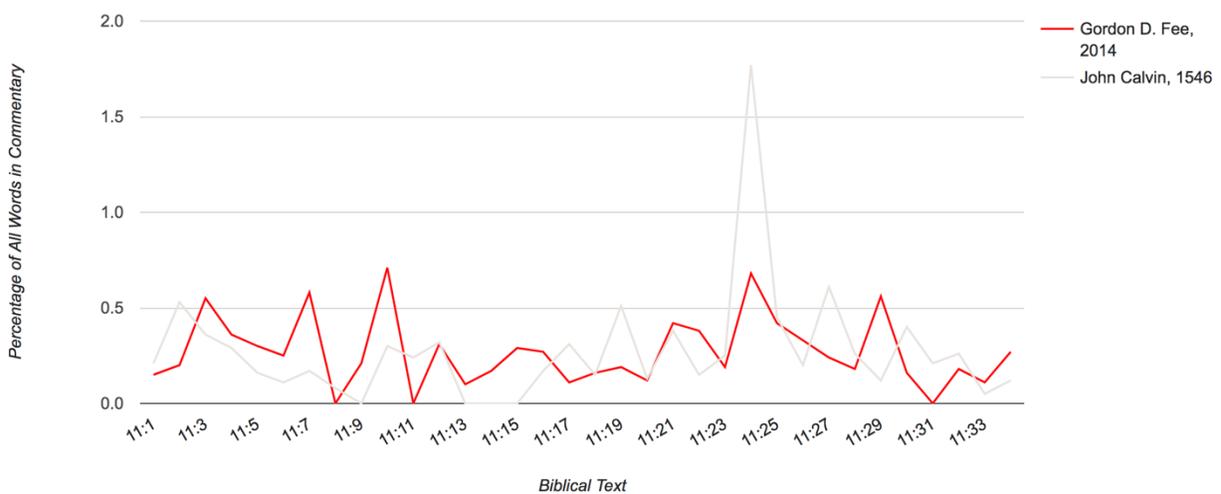


Figure 19: Comparison of the Relative Treatment of Each Verse in 1 Corinthians 11, Fee and Calvin

The interpreter now has a place to begin her comparison between the two commentaries. Taking 11:24 as the lens through which to compare, one can begin to see that the theological debates of Calvin's day, in this case the debate over the taking of the Lord's Supper, Paul's particular focus in 1 Corinthians 11:24, drives his analysis.⁴ In comparison, Fee pays little attention to the traditionally theological questions of 1 Corinthians, and instead is much more interested in questions related to the 1st-century setting of Paul's letter. Consider, for example, the relatively extensive treatment Fee gives to Paul's reference to slavery in 7:21 (0.64% of the total words in Fee's commentary) in comparison to Calvin (0.19% of the total words in Calvin's commentary).⁵ These initial insights, gained by the analysis of the words at a distance, become the grounds for close reading and careful comparison of the two approaches to the letter.

In this chapter, I outline the process by which one could expand this type of analysis over many more commentaries to better understand the development of the genre. As with the approach outlined in the previous chapter, using OCLC metadata to understand the changes in the commentary genre, this approach is not offered as a replacement for close reading. Rather, by considering the relative treatment of particular verses and chapters, readers may more easily find distinctive commentaries and, when combined with publication metadata, better understand the development of this genre and the development of the reading of the New Testament. Is Calvin's focus on theological issues and Fee's focus on historical issues consistent with other commentaries of their respective eras? The data set is too large to look at this question through

⁴ 1 Corinthians 11:24 (NRSV): "And when he had given thanks, he broke it and said, 'This is my body that is for you. Do this in remembrance of me.'"

⁵ 1 Corinthians 7:21 (NRSV): "Were you a slave when called? Do not be concerned about it. Even if you can gain your freedom, make use of your present condition now more than ever."

close reading. I suggest statistical analysis of relative treatments is a more effective way to answer the question.

In this chapter I begin by describing the process of digitizing, tagging, and parsing full-text commentaries necessary for this type of analysis. That is, just like the previous chapter where I first introduced a new data set (OCLC metadata) before discussing the distant reading approach to that data set, I begin here by introducing the challenge of working with the digitized corpus of digitized commentaries. I then turn to demonstrate the process of distant reading by analyzing a set of digitized commentaries on the New Testament letter of James. I then end by suggesting how New Testament scholars and librarians can work together to scale this approach to help better understand how the commentary genre has developed over time.

Using the Commentary Form as Structured Textual Data for Distant Analysis

The goal for the type of distance analysis proposed in this chapter is to intuit the relative importance a given commentator places on particular New Testament texts by determining the relative length of treatment in a commentary. The process for this analysis is rather straightforward: divide a given commentary into specific chapters and verses of the treatment of a New Testament text, write software to count the number of words dedicated to the treatment of each section of the New Testament text, and then map changes in this relative treatment over time. In this section I outline the steps necessary to conduct word counts of sections of New Testament commentaries.

Locating Digitized Commentaries

The obvious starting point for distant reading commentaries is locating digitized commentaries ready for analysis. To perform distant reading that allows for substantive conclusions about the development of the genre of New Testament commentary, readers need a large number of commentaries in digital form, with accurate OCR for word counts. This initial step in the analysis of New Testament commentaries proves to be more difficult than one might imagine. When reading the works of pioneers of distant reading like Moretti and Jockers, one is left with the impression that these scholars are spoiled with a plethora of digitized texts for analysis. Jockers' analysis of nineteenth-century novels, for example, drew on an existing digitized corpus of 3,346 books.⁶ For other examples of distant reading, specialized databases with digitized collections make locating large selections of digitized texts for distant reading a quick step in the analysis.⁷ With other examples of distant reading, bibliographic metadata is of primary interest, and thus the digitized corpus is much more easily attainable.⁸ The rise of mass digitization of books is making available millions of texts with character recognition. Google Books, HathiTrust, and the Internet Archive are just a few examples of the many sites that are collecting digitized forms of books in many genres, and thus seemingly creating the opportunity for distant reading.⁹

⁶ Matthew Jockers, "Theme," in *Macroanalysis* (Urbana: University of Illinois Press, 2013), 118–53.

⁷ Consider, for example, Dallas Liddle's use of the *Times Digital Archive*, which provides access to over 20,000 digitized Victorian newspapers: Dallas Liddle, "Reflections on 20,000 Victorian Newspapers: 'Distant Reading' The Times Using The Times Digital Archive," *Journal of Victorian Culture* (Routledge) 17.2 (2012): 230–37.

⁸ See, for example, Moretti's work with only the titles of 7,000 British novels: Franco Moretti, "Style, Inc. Reflections on Seven Thousand Titles (British Novels, 1740–1850)," *Critical Inquiry* 36.1 (2009): 134–58.

⁹ It is important to recognize that even these large collections of digitized texts are collaborative efforts, and thus there is no systematic quality control. The OCR in many 19th-century texts, for

For the analysis of New Testament commentaries, though, a large, digitized corpus is difficult to find. There is a surprisingly small number of digitized texts in these standard online collections. For example, Internet Archive contains only 272 commentaries on Paul's letter to the Galatians. HathiTrust contains only 45 full-text commentaries on Galatians. Even more problematic, the available selection is not representative of the entire tradition at all. The Internet Archive commentaries on Galatians are overwhelmingly in English (260 out of 272). Almost all are in the public domain, with only one available that was published after 1923. Jockers' joke that "Today's digital-minded literary scholar is shackled in time; we are all, or all soon to become, nineteenth centuryists," seems certainly to be true for those interested in distant reading New Testament commentaries.¹⁰ This available selection of digitized commentaries is idiosyncratic and subjective, as the makeup of the library is dependent upon the opaque selection criteria of contributing institutions.

To perform the distant reading of the genre I introduce here, therefore, one must first seek out alternative methods of accessing digitized commentaries. Simply using what is available online will not suffice. For this study, I propose two methods of locating a better selection of texts. The first is to digitize newly commentaries from eras that are missing from the publicly-available corpora.¹¹ With the lower cost and greater availability of book scanners and high-

example, proves to be almost unusable. For a discussion of quality control in HathiTrust, see <https://www.hathitrust.org/quality-in-hathitrust>.

¹⁰ Matthew Lee Jockers, *Macroanalysis Digital Methods and Literary History*, Topics in the digital humanities (Urbana: University of Illinois Press, 2013), 173.

¹¹ For a discussion of how one should undertake a digitization project, see Ray Abruzzi, "Undertaking a Digitization Project?," *Online Searcher* 39.2 (2015): 28–32.

efficiency document feeders, this is becoming more possible than it was in the past.¹² Scholars and librarians can grow the corpus by digitizing print commentaries. There are, however, several challenges with the approach of digitizing print libraries for analysis. The first, of course, is the challenge of scale. No individual can, with any expectation of efficiency (or sanity!), take on the process of growing significantly the corpus of online commentaries. The size of the corpus demands collaboration between libraries and archives, and professional organizations may play an important role in coordinating these efforts.¹³ This problem of scale is compounded by the fact that digitization efforts require more than simply scanning materials. Scanning must be accompanied by the creation of accurate and thorough metadata and OCR correction, time-consuming steps often lost on patrons who demand texts should be available online.¹⁴

This need to grow the insufficient corpus of digitized commentaries leads to a second challenge: working with copyright restrictions on non-public domain works. As noted, most of the publicly-available digitized New Testament commentaries were published before 1923 and thus in the public domain. There is great need to add digitized copies of in-copyright commentaries for analysis, and if one is going to the effort to digitize works that are in copyright or orphaned works for a distant reading project, then one would like to have the ability to share this work with others. However, it is best practice to avoid sharing works that are in-copyright or those for which the copyright status is not clear. For significant distant reading to occur on New

¹² For a discussion of the increasing affordability of high-end scanning equipment, see Yongli Zhou, “Fulfill Your Digital Preservation Goals with a Budget Studio,” *Information Technology & Libraries* 35.1 (2016): 26–50.

¹³ For a promising example of collaboration, see the American Theological Library Association’s Digital Library (<http://dl.atla.com>).

¹⁴ For a clear argument about the challenges of digitizing materials, particularly archival materials, see Larisa K. Miller, “All Text Considered: A Perspective on Mass Digitizing and Archival Processing,” *The American Archivist* 76.2 (2013): 521–41.

Testament commentaries there must be significant efforts to digitize commentaries, which means that scholars must understand copyright law. Here again there is an opportunity for librarians and scholars to work together to navigate these difficult waters.

Distant reading these works through the method proposed here is an example of non-consumptive use of digitized texts.¹⁵ In the mind of some, however, it is still a matter of unsettled law as to whether one can even legally mine digitized texts if those texts are still under copyright.¹⁶ Therefore, even as digitization efforts are made by individuals for distant reading purposes, whether on their own or through the procurement of publishers' digital copies, readers will not be able to distribute their digitized texts.

The challenge of copyright leads to a second possibility for growing the available corpus of digitized commentaries for distant reading: working with copyright holders. With the growth in e-book platforms and digital publishing, more print books exist in clean, digital forms, particularly recent publications. Scholars interested in mining these texts may need to work with publishers and authors to attain digital texts for analysis. In the case of Biblical studies, many of these texts are available through commercial Bible software providers like Accordance, Bibleworks, and Logos, as well as other specialized Bible websites.¹⁷ While publishers in religious studies are not as up to date in terms of offering electronic versions for purchase on an individual basis, they are working with these software providers to provide digital texts. For

¹⁵ For a full discussion of non consumptive use of digitized, in-copyright texts, arguing for the legality of such practice, see Matthew Sag, "Orphan Works as Grist for the Data Mill Symposium: Orphan Works & Mass Digitization: Obstacles & Opportunities," *Berkeley Tech. L.J.* 27 (2012): 1503–50.

¹⁶ Berkeley Law hosted a 2012 conference on this issue. For notes from the panel discussions, see <https://www.law.berkeley.edu/research/bclt/past-events/2012-conferences-2/april012-orphan-works-symposium/agenda-ppts-audio/>.

¹⁷ See <http://www.accordancebible.com>; <http://www.bibleworks.com>; <http://www.logos.com>.

example, the well-regarded Yale Anchor Bible series of commentaries is not available in any of the major e-book platforms, but all commentaries in the Yale Anchor Bible series are available for purchase as add-ins to the Logos software suite. This is a promising source of digitized commentaries, though offerings are limited and in many cases expensive.

This challenge of copyright is connected to a third, broader challenge with digitizing New Testament commentaries: the mechanism by which candidates for digitization are chosen. As noted in the consideration of the digitized commentaries that are available, there is a necessary subjectivity to the process of digitization. In the absence of a completely digitized corpus of commentaries, which seems impossible given the size of the corpus highlighted in the previous chapter, those engaged in distant reading will be at the mercy of what is available or what they decide to digitize for themselves. A necessary question for any distant reading, therefore, is what commentaries to digitize and analyze? Certainly distant readers will want to cultivate a diversity of authors, time periods, and perspectives in any corpus of analysis. Therefore, a consideration of bibliographic metadata will be important for considering the corpus to use for distant reading. In fact, analysis like the type proposed in the previous chapter may be helpful as a prior step to the type of analysis proposed in this chapter. If one can detect inflection points in the publication history of the commentary genre, one may be able to select appropriate places for digitization.

For the demonstration of the distant reading method below, I have selected commentaries on the letter of James and gained access to these commentaries through a number of methods, demonstrating some of the proposed options above, including some harvested from the Internet Archive, some digitized myself for analysis, and some purchased from commercial Bible software providers. My dataset, as described below, was not chosen as a representative sample of all commentaries on James, but rather, as a way of demonstrating these methods of digitization.

If interpreters are going to engage in more thorough distant reading, though, the availability of digitized commentaries must improve.

Encoding Commentaries for Analysis

Gaining access to digitized commentaries with character recognition is just the first challenge in finding texts ready for the type of distant reading proposed here. I am proposing that a helpful way of studying the development of this genre is by looking at the relative size of treatment given to each verse or chapter of the New Testament. To perform this analysis, one needs not only a digitized corpus of commentaries, but a corpus of texts divided into verse and chapter sections that a computer can read. That is, while the pages of individual commentaries present visual markers identifying where the commentator moves from one verse to another, those markers must be standardized across a digital corpus, and they must be readable by software, rather than humans.

Fortunately for distant readers, there have been great developments in textual markup in the past few years. The standardization of formats like XML and JSON, along with editors and tools for creating and modifying XML, make the organization of digital texts into machine-readable formats relatively easy.¹⁸ As more publishers release digital texts formatted in standards like XML, distant readers will be able to more easily write software to mine patterns in texts.¹⁹

¹⁸ There are many introductions to the growth of XML in textual analysis. For representative samples, see Juliet L. Hardesty, “Transitioning from XML to RDF: Considerations for an Effective Move Towards Linked Data and the Semantic Web,” *Information Technology & Libraries* 35.1 (2016): 51–64. Adam Zukowski, “XML for Catalogers and Metadata Librarians,” *Library Quarterly* 85.3 (2015): 331–33. Airi Salminen, Eliisa Jauhiainen, and Reija Nurmekkala, “A Life Cycle Model of XML Documents,” *Journal of the Association for Information Science & Technology* 65.12 (2014): 2564–80.

¹⁹ A critical development in this area is the growth in the use of the Text Encoding Initiative (TEI), a set of standards for encoding texts to be represented in a variety of formats. This

Unfortunately for the purposes of tracing commentaries' treatment of given sections of a New Testament text, content-specific XML encoding is necessary. That is, while XML schema like the Text Encoding Initiative (TEI) are based on formatting and display, the method proposed here depends upon an outline of a text based on content. Therefore, the tagging of these commentaries must be done by hand, with someone identifying where the treatment of one verse ends and another begins.²⁰ A second step in the analysis, therefore, is to encode full text commentaries with nested XML, indicating the sections of a commentary relevant to a given chapter or verse of the New Testament text. In the demonstration of this method, I have encoded the full-text commentaries with specialized XML tags to divide a commentary into the relevant sections of the New Testament.²¹ These XML tags, which label section of the commentary by chapter and verse number, allows software to conduct word counts of relevant sections.

One assumption has been made in the tagging of these texts. Often a commentator will not comment merely on a single verse, but rather on a cluster of verses together, without clearly indicating where the treatment of one ends and another begins. A consistent example in James

standard XML protocol allows software to “understand” sections of a text, including formatting elements like paragraphs and grammatical structures. For an introduction, see <http://www.tei-c.org/>. For a discussion of the role of librarians and TEI, see Lindsey A.1 Halsell lindseyhalsell@gmail.com, “TEI and So Can You: Corpus Linguistics and the Text Encoding Initiative,” *PNLA Quarterly* 77.3 (2013): 63–70. For an example of the effective use of TEI, see Sharon Rankin and Casey Lees, “The McGill Library Chapbook Project: A Case Study in TEI Encoding,” *OCLC Systems & Services* 31.3 (2015): 134–43.

²⁰ It was one hope of this study to identify a more automated mode of tagging based on content. Certain print commentaries have textual clues that indicate a shift in verse or chapter. However, this is not standard, and therefore it cannot be universally applied. Alternatively, one could physically mark an item before digitization to indicate where these breaks are, but such marks can upset the optical character recognition. One promising lead is that many Bible software applications key commentaries to the Biblical text, and therefore they are already encoding their texts by verse and chapter. However, these standards are idiosyncratic to the software system, and they are mixed with other tags that upset XML parsing. Below I discuss the possibility of standardizing a format for encoding commentaries and other works of Biblical interpretation.

²¹ An example of an encoded section of one commentary is available as Appendix 3.

commentaries is James 5:19-20, the closing comments of the letter, which are generally treated as a single unit. In my XML tagging, I place the tags next to one another for verses that are treated together, such that the software will tally the first verse (in the example, 5:19) as having zero words of treatment and the second verse as having the full treatment. In the analysis of results, therefore, readers should understand that a verse that receives zero words is a verse that is covered in the subsequent verse's section.

Verse by Verse Analysis of Commentaries

Once a reader has obtained a set of digitized commentaries, encoded to indicate which sections of the New Testament text are covered, software must parse the texts to count the number of words in each section. For this application, I have written a Python script that parses through a full-text commentary and, based on the custom XML tags, counts the number of words focused on a given chapter and a given verse.²² The script then stores these values in a SQL database, so that the analysis of the commentaries can take place offline, not depending upon continual parsing of the texts to draw conclusions.

Parsing these texts based on the XML encoding provides a rough dataset upon which to analyze differences in the commentaries. Visualization is the most helpful way to discern changes or patterns, and so I have written a visualization application that draws upon the SQL database and uses the Google Maps API to map these commentaries' relative treatment of chapters and verses. To make comparison most productive, I display each commentary based not on how many words it gives to a given verse, but rather what percentage of the commentary's

²² All software is available at <https://github.com/boadams01>.

total word count is dedicated to this verse. This helps to control the wide diversity of lengths of these commentaries.

A Small Demonstration of Distant Reading Commentaries as Structured Textual Data, Using the Letter of James

With this basic introduction to the method in place, I turn to demonstrate the potential of the method with a small corpus of commentaries on a single book of the New Testament. For this demonstration, I have selected 8 commentaries on the letter of James, a 5-chapter letter toward the end of the New Testament. The power of distant reading is the ability to process large corpora of texts and detect patterns, avoiding the idiosyncrasies of a given example. Therefore, this corpus is far too small to make conclusive claims about the development of the commentary genre. However, analysis with this small corpus will demonstrate how reading across commentaries through computational analysis can be more productive than close reading representative samples. After discussing these results, I will consider how this project could be scaled so as to allow an interpreter to make more significant claims.

Definition of the Corpus

These texts selected for this demonstration range in publication date from 1550 (Calvin) through 2000 (Moo), with most published in the 19th and 20th centuries. These works were chosen not to be representative samples of the genre as a whole, but rather to demonstrate a range of digitization techniques. The works range in size from 202,215 words (Martin) to 17,495

words (Reicke), and they were digitized for this study by a number of different methods.²³ The corpus as a whole consists of 758,451 total words, with 40,293 unique word forms.

Author	Publication Year	#Words²⁴	Method of Digitization
Calvin	1550	31,661	Purchased from Accordance
Davids	1983	71,773	Purchased from Accordance
Henry	1737	33,697	Download from BibleStudyTools.com
Manton	1647	176,296	Purchased from Accordance
Martin	1988	202,215	Self Scanned
Moo	2000	94,425	Purchased from Accordance
Reicke	1964	17,527	Self Scanned
Ropes	1916	129,857	Download from HathiTrust

Figure 20: The 8 Commentaries in the Corpus Studied

Not surprisingly, the words “God” and “James” dominate the corpus, appearing more than twice as many times as any other word.²⁵

²³ See Appendix 2 for a full bibliography of these 8 commentaries.

²⁴ Word counts are not representative of the entire commentary, but rather the section of the work that is closely connected to the exposition of the Biblical book. Series prefaces, abbreviation lists, indexes, and bibliographies have been removed for the purposes of word counts.

²⁵ The word “James” appears 4,945 times (0.65% of words) in the corpus and the word “God” appears 4,176 times (0.55%). The word cloud and other content statistics were generated using the Voyant online text analysis tool (<http://voyant-tools.org>).

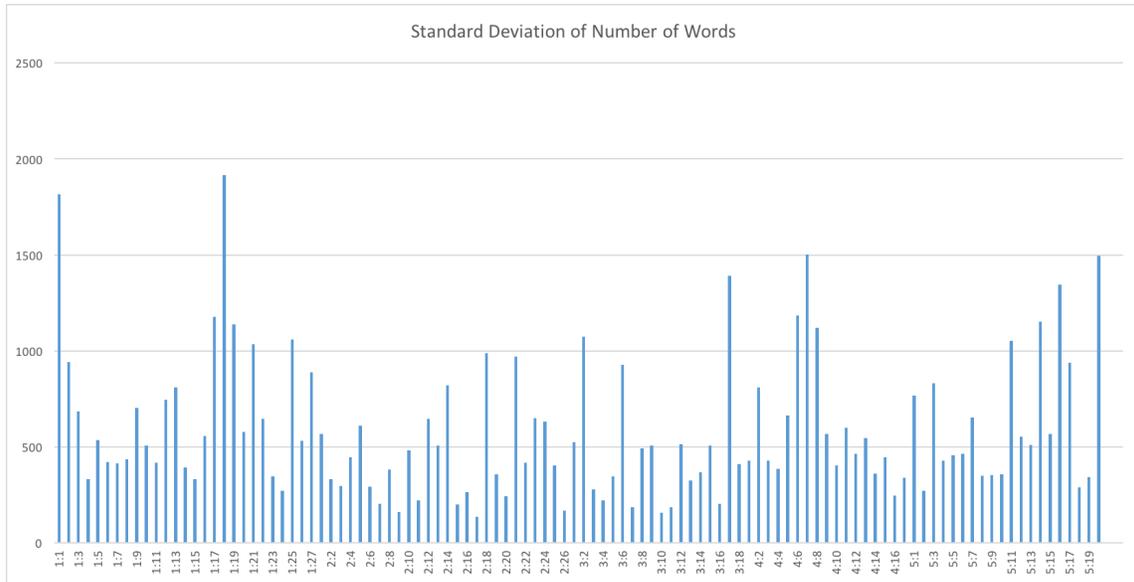


Figure 22: Standard Deviations of Verse Treatments

This graph itself can be helpful for approaching the set of commentaries, as it suggests that verses like 1:18 or 4:7 may be productive places to discern changes in the tradition, as the interpreters differ widely in how much attention they devote to these texts.

Though this set of commentaries is small in comparison with the total number of commentaries published on James, interesting insights result from stepping back and reading the commentaries from a distance. I here offer a few small insights on the development of James scholarship compiled through the analysis of commentaries outlined above. These insights are offered not as definitive statements about the reception history of this text, but rather as demonstration of the potential power of this method.

Tracing Chronological Shifts from Theology to History Based on the Most Significant Texts

One potential insight from distant reading is the ability to discern interpretive strategies of a commentary based on the amount of focus paid to particular parts of Scripture. If it is safe to assume that an author's interpretive strategy can be identified where he or she commits more

words, then the interpretive priorities of commentaries can be discerned based on what types of verses receive the most attention. Alternatively, verses that receive more significant treatment may be an indication of topics that were of interest in the scholarly tradition or in the culture more broadly, as we saw with Calvin's focus on 1 Corinthians 11:24. When combining this with publication metadata like publication date, then interpreters can begin to understand how interpretive priorities have changed over time.

In the case of these commentaries on James, based on the relative attention given to particular texts, readers can detect a pronounced growing interest in historical questions in recent commentators, as opposed to the theological ones that interested older commentators. Consider, for example, a comparison of relative treatment between the commentary of Thomas Manton (1647) and Douglas Moo (2000). The verses that attract relatively more attention from Moo are those that reference aspects of life in the letter's author's first century culture. Moo dedicates, for example, 1,591 words (1.6% of the commentary) to the small verse 3:12 that draws upon nature imagery of the first century, whereas Manton only offers a comment of 194 words (0.11% of his commentary).²⁶ By contrast, verses that attract relatively more attention from Manton are those that focus on matters theological, and Moo pays them little attention. Consider, for example, that Manton uses 5,822 (3.34% of the commentary) words to comment on James 1:18, a verse that identifies the creation of humans as the first fruits of God's creation, whereas Moo's comment is a far briefer 604-word note (0.16% of the commentary). Manton's focus on theology seems consistent with those who came before him. John Calvin (1550) likewise focuses much attention on 1:18 (448 words or 1.42%) and almost ignores 3:12 (59 words or 0.19%). It seems, therefore,

²⁶ James 3:12: Can a fig tree, my brothers and sisters, yield olives, or a grapevine figs? No more can salt water yield fresh (NRSV).

that there may be a chronological shift away from focusing on a more theological verse to focusing on one related to imagery from the first century. Such a suggestion seems to be confirmed by looking at the early interpreter Matthew Henry, who, like Calvin, pays little attention to 3:12 (47 words or 0.14%). The chart showing the percentage of words written on 3:12 in all 8 commentaries seems to confirm this trend (Figure 7).

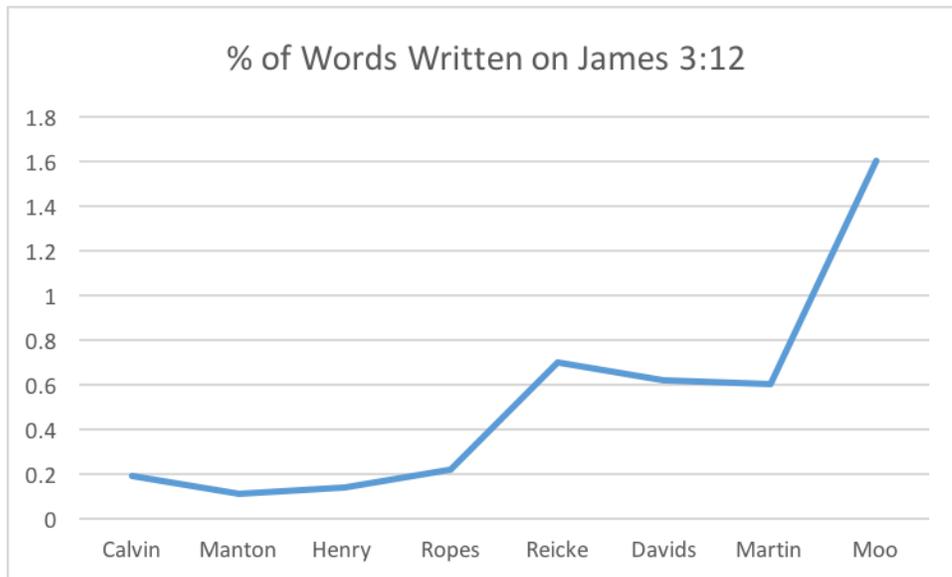


Figure 23: Percentage of Words Written on 3:12, From Earliest to Most Recent

This chronological shift of interpretive focus from theology to history, of course, will not be surprising to anyone familiar with the history of Biblical interpretation. Indeed, entire histories of Biblical studies have been written arguing that in the late-18th and early-19th centuries, scholars began viewing the New Testament as a window into history rather than a theological tractate. The most prominent example of an argument for this shift is the work of Hans Frei.²⁷ The promise of this method of distant reading, though, is that analysis can identify individual commentaries that do not fit the trend, a trend interpreters like Frei identify based on

²⁷ Hans W. Frei, *The Eclipse of Biblical Narrative: A Study in Eighteenth and Nineteenth Century Hermeneutics* (New Haven: Yale University Press, 1974).

the selection of individual examples deemed representative of a chronological period. Reading more commentaries at a distance can more easily reveal outliers of established trends. For example, Moo's rough contemporary Ralph Martin (1988) dedicates 3,057 words (2.92% of his commentary) to 1:18-19, a verse that is explicitly theological. It would be worth considering, therefore, whether Martin's commentary is more interested in matters theological than Frei's model would suggest he would be based on the time period of his writing.

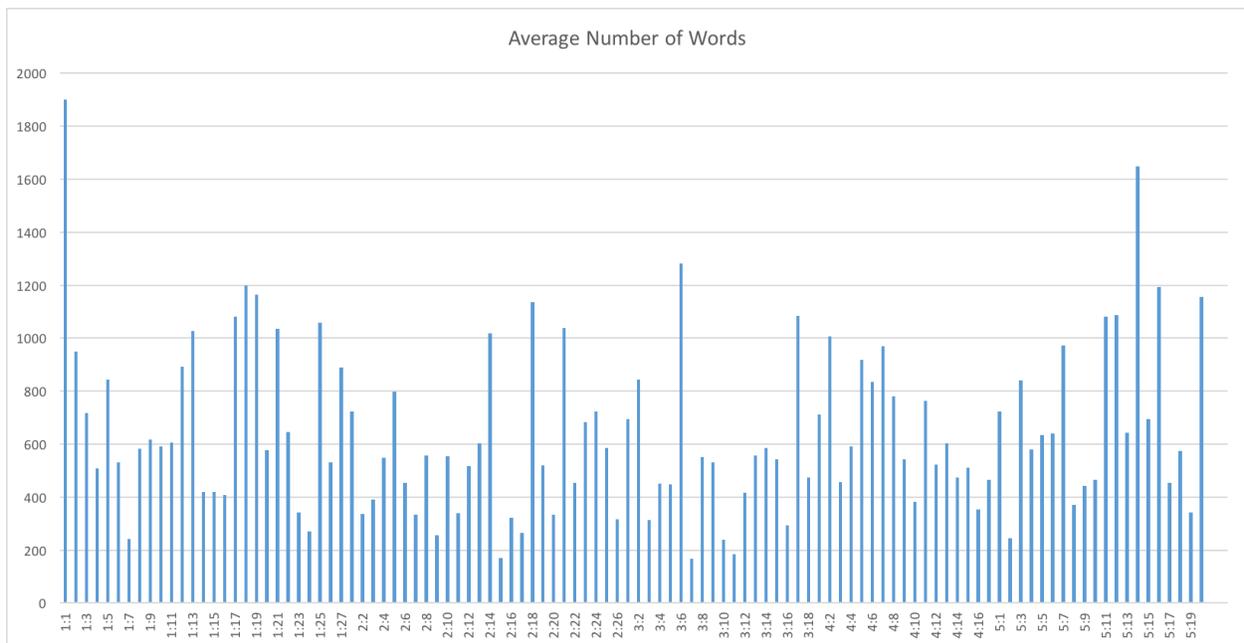
The intention here is not to make sweeping generalizations about the history of commentaries on James. The intention is, though, to suggest that any generalizations are safer when based on analysis of more examples of the genre than is typical in works like Frei. Reading at a distance through computational analysis, as demonstrated here, allows interpreters to see more examples of scholarship at different periods of time. An interpreter reading from a distance is more likely to find the commentary that resists the generalized narrative of the development of the commentary genre. The close reader like Frei must necessarily sacrifice breadth of inquiry in order to focus on content; representative samples must be selected. The distant reader, however, sacrifices the focus on content through close reading, in this case using length of treatment as an assumption of focus, in order to consider more examples. Distant reading of the commentary genre, therefore, may be a helpful corrective to narratives constructed by subjective choices of representative samples.

Comparing Commentaries and the Periodical Literature on James

Distance analysis on commentaries may also open up new avenues of research into the scholarly tendencies of Biblical scholars and the ways in which genres of scholarship affect the content of scholarship. Do scholars writing commentaries focus on the same issues that they

focus on when writing journal articles? The type of analysis of the commentary genre may allow readers to begin answering this and similar questions. Periodical literature indexes have existed for decades, and they allow a more granular view of what scholars are publishing on. For example, journals in Biblical studies have been indexed by Scripture by the American Theological Library Association. It is possible to identify journal articles written on specific passages of the New Testament.²⁸

Combining this existing search capability with the distance analysis of commentaries outlined in this chapter provides a more detailed picture of what is interesting scholars and how those interests have changed over time. In analyzing this small set of 8 commentaries on James, I have identified the verses that receive the heaviest treatment (see Figure 6) in the letter.



The opening verse of James receives the heaviest treatment (an average of 2353.5 words), which is to be expected, given the tendency of scholars to comment on aspects of the rest of the letter when commenting on a salutation. Heavy treatment also falls on 1:18 (1528.5 words), 3:6 (1602.3 words), and 5:14 (1960 words). Given the nature of these verses, this may not be surprising to readers. As noted, 1:18 is one of the more significant theological statements of the letter, 3:6 is a famous statement about the tongue setting the entire body ablaze, and 5:14 is a practical instruction about elders praying over the sick, an idea that has received much attention in the history of theology. As noted, though, 1:18 is treated extensively by some early commentators, but not as much by more recent commentators. This variation in treatment is evident in the high standard deviation. In fact, 1:18 has the highest standard deviation (2150.57) of all verses of James in the set of commentaries considered. The standard deviations for 3:6 (841.29) and 5:14 (1172) are far lower.

Comparing these results to searches in the Scripture index of the ATLA Religion Database (see Figure 8), we find slightly some similar but also some very different results.

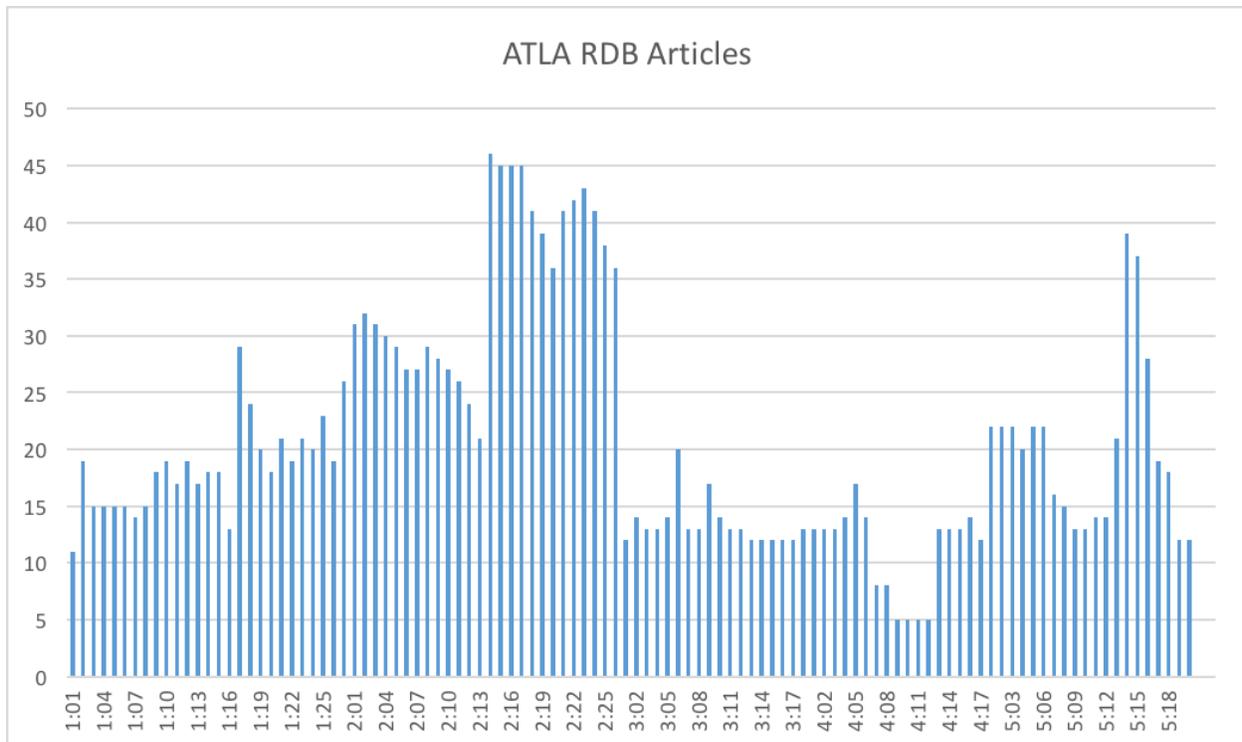


Figure 25: Number of Articles Index by Verse of James in the ATLA Religion Database

James 5:14 receives a large amount of attention in the periodical literature, just as the distant reading of the commentaries showed. By comparison, though, 3:6 receives far less attention in the periodical literature than in these commentaries. More popular in the periodical literature is the passage 2:16-26, the famous discussion of faith and works in James that seems to be at odds with Paul’s discussion of this same topic in Romans and Galatians. It may be that the commentary genre does not afford interpreter the space to discuss the complexity of this passage within its New Testament context. The distinction in treatment between the commentaries and the periodical literature is likely a place of fruitful engagement.

These two brief examples show the potential for this type of distance analysis. In each case, the focus is not on the content of the argument of the commentary, but rather on seeing individual priorities within a commentary or mapping how different parts of the Biblical text is

handled across commentaries. The benefit of the process outlined above is that once the commentaries are encoded, analysis can proceed in many different directions.

Expanding and Scaling this Distant Reading Approach

This demonstration explores only a few commentaries, and thus it does not take advantage of the true power of distant reading, where the enormous size of the corpus is used to the advantage of the reader. The comments above about developments of the reading of James likely fall victim to the idiosyncrasies of individual commentators, and thus they represent only a slight improvement beyond the representative sample method evident in previous close reading analysis, like that of Childs introduced in the opening chapter of this study. To harness fully the power of Moretti's vision for distant reading, interpreters would need to work on a much larger sample of commentaries on James.

However, the method here is easily scaled. The parsing software, database structure, and visualization engine deployed on 8 commentaries could easily be deployed on 800 commentaries, with no changes. This would require, of course, that the interpreter had access to 800 commentaries with proper XML encoding; this is no small task! Below I identify two steps that librarians and Biblical scholars can take together to facilitate this type of scaling of this approach.

Librarians Should Work to Increase the Availability of Digitized Commentaries

As noted above, the corpus of Biblical commentaries available digitally, with clean OCR, is surprisingly small. This is all the more surprising given the ease with which institutions can digitize their collections and the multiple repositories available to store texts and make them

discoverable. There should be, therefore, a clear effort by all institutions to do their part to fill in gaps in what is available.

However, digitization efforts should be coordinated to achieve maximum efficiencies. It would be most helpful for large institutions that have the resources to digitize collections to work together to ensure that there is no duplication of efforts. Therefore, librarians must work with their counterparts at other institutions to create a strategy for digitization. One clear lesson from this study is the need for libraries to digitize their collections of Biblical commentaries and make them available either in institutional repositories or in centralized repositories like HathiTrust. These digitization efforts should be coordinated, and they should be done with care to ensure that there is proper OCR and proper metadata, using up to date cataloging standards, including linked data.

As noted, however, even the best efforts of digitization will fall victim to the limits of copyright. Post-1923 commentaries cannot be made available online. However, I would feel safe publishing the results of non-consumptive parsing of these post-1923 commentaries (see Appendix 1). This data may help scholars who are interested in this type of analysis. Therefore, a second digitization effort would include librarians working with scholars to make available datasets that result from distant reading. The digital age offers great opportunities for collaboration, even when that collaboration happens anonymously through the publication of research data. This study highlights the importance of sharing digitized texts and distant reading data to help scale this method to make significant contributions to understanding the history of Biblical interpretation.

Librarians Should Approach Publishers and Scholars about Alternative Forms of Publication

This study has highlighted the importance of having access to texts formatted for suitable machine reading. Increasingly publishers are using XML-encoded formats, including TEI, to format and distribute their works. This effort should not only aid in the formatting of a work for printing or displaying, but it could also be used for reading. Therefore, librarians should take the initiative to seek out these alternative forms of publication, particularly if they are already purchasing the print form of works. The challenge of distant reading, as I have noted, is that it represents a shift in the fundamental paradigm of the scholarly process. This shift is also hard to swallow for publishers, who are used to producing work in the most human-readable form possible. However, as machines take up more of a role in helping the scholar read, publishers should be interested in making available these alternative forms of works for scholarly consumption.

In addition to contacting publishers, librarians should work with individuals who are producing scholarship to suggest they consider encoding their own work. Any librarian who has worked with the writer of a commentary knows that these works are, in the workspace of the scholar, already organized into discrete sections of the Biblical text. An effective librarian would show the scholar how to encode that organization into XML format, to more effectively share his or her work with colleagues across the field.

Librarians are uniquely positioned between scholars and publishers, and therefore librarians must take a leading role in expanding the notion of reading and publishing with which these actors operate. The scalability of distant reading depends upon librarians working to increase access to encoded texts.

Librarians Should Work to Standardize Encoding Formats

As indicated above, there has been helpful standardization of XML formats that allow for publishers, print and digital, to create texts that are more portable. These standards also allow distant readers to better understand shifting characteristics of large corpora of texts.²⁹ For the type of distant reading introduced in this section, though, TEI is not too helpful. What distant readers of the scholarship of the Bible need is an encoding format that is specific to the Biblical text. Because the field of Biblical studies is focused around a singular (and relatively small) set of texts, it would make sense for librarians to determine an XML scheme by which scholarship, be it monograph, commentary, or journal article, could be encoded to allow for new distant methods. Once again, librarians are uniquely positioned to begin this conversation with scholars and publishers, and therefore they should take the lead.

It has been the goal of this chapter to introduce a method that is scalable. As noted, though, the promise of this scaling depends upon the increased availability of digital texts, available in formats that distant readers can use. I offer a challenge to librarians to work with publishers and producers of scholarship to not only advertise the power of distant reading like the method introduced here, but also to work to make the texts available to grow this method.

²⁹ For a good introduction and discussion of how TEI can be used in distant reading, see Michael Widner, “Text-Mining the Middle Ages | Manuscripts and Machines,” January 25, 2016, <https://people.stanford.edu/widner/content/text-mining-middle-ages>.

CHAPTER 4: THE PROMISE OF DISTANT READING IN BIBLICAL STUDIES

“We have not yet seen the scaling of our scholarly questions in accordance with the massive scaling of digital content that is now held in twenty-first-century digital libraries.” –Matthew Jockers¹

Matthew Jockers’s prediction about the future of humanities and textual scholarship captures the hope that many of us librarians have for shifts in modes of scholarship. The growth in the digitization of the products of scholarship opens up possibilities for using machines to understand the changes in the history of scholarship. New texts and new tools present opportunities for librarians to work with subject matter experts to understand better how specific topics have been studied and to discover new insights in the textual tradition.

Jockers’ prediction, though, is also a reminder that many fields of scholarship are not there yet. Many in the humanities are operating with the same modes of reading with which their predecessors have operated for decades or, in the case of New Testament studies, centuries. The digitization of texts and development of text mining tools remain for many a mere novelty, interesting ideas that are treated as separate from textual scholarship. Jockers and his colleagues at the forefront of distant reading decry this attitude not simply as a missed opportunity. Rather, they argue that to ignore the potential of mining large corpora of texts is to continue to fall into

¹ Matthew L. Jockers, *Macroanalysis: Digital Methods and Literary History*, 1st Edition. (Urbana: University of Illinois Press, 2013).

the trap that textual scholarship has long operated in, reading a large, diverse scholarly tradition through the lens of a select few, “canonical” texts.

Those who work in Biblical studies are acutely aware of this danger of the traditional form of textual scholarship. For centuries, the conversation in Biblical studies has been dominated by the voices of a few, select leaders, and those leaders most often are a steady cast of white, European males.² The rise of alternative voices in Biblical scholarship is a welcome phenomenon, one far too late in arriving, that enriches the conversation for all.³

It has been the goal of this study to provide another form of reinvigorating this tradition of scholarship, aimed likewise at enriching the conversation. Distant reading of Biblical commentaries, whether through analysis of OCLC metadata or by tracing the relative importance of particular passages of Scripture, provides a new lens for scholars to see the long history of interpretation of the text. This distant reading should operate in coordination with the close reading methods so familiar to Biblical scholars. By using digitized texts and custom software to scan across the enormous corpus, scholars may discover trends, patterns, and new works that may enrich their understanding of the how reading the Bible has changed over time and who has been responsible for this change.

Such analysis is only possible, though, if the tools for distant reading are made available. And therefore, I hope this study is as much a call to action as it is a description of method.

² For an important critique of the field of New Testament studies along these lines, see Brian K. Blount, *Cultural Interpretation: Reorienting New Testament Criticism* (Minneapolis, MN: Fortress Press, 1995).

³ For examples of the expanding conversation in Biblical commentary writing, see Carol A. Newsom, Sharon H. Ringe, and Jacqueline E. Lapsley, eds., *Women’s Bible Commentary, Third Edition: Revised and Updated*, 3rd Anniv ed. edition. (Louisville, Ky: Westminster John Knox Press, 2012); Brian K. Blount et al., eds., *True to Our Native Land: An African American New Testament Commentary*, 5.2.2007 edition. (Minneapolis: Fortress Press, 2007).

Jockers' "massive scaling" overstates the current state of digitization in Biblical studies, and so librarians must work with scholars and publishers to increase ready access to the digital corpus of Biblical studies.

This study began by identifying this need for distant reading in Biblical studies, a method where the enormous corpus of Biblical commentaries is seen not as a burden, hindering scholarship, but rather as a benefit. The large data set potentially allows scholars to see shifts and changes that define the genre and field, rather than those that define individual works. In chapters 2 and 3, I have proposed two specific methods for mining this large corpus of New Testament scholarship. In the first, I introduced bibliographic metadata as focus of study, one that is likely new to Biblical scholars. Technological advances such as the OCLC WorldCat API make accessible this vital information that can supplement the content of Biblical commentaries to indicate trends in the study of the New Testament. In the third chapter, I introduced a method for distant reading commentaries themselves, through topic modeling based on chapter and verse. Biblical scholars are beneficiaries of a stable form of scholarship, and this method allows scholars to use that form to their advantage.

These two methods, however, certainly do not exhaust the possibilities for distant reading in Biblical studies. Dozens of other distance approaches would be equally feasible, allowing scholars to take advantage of this large, but consistent corpus of literature to understanding the development of the practice of Biblical interpretation. For example, one could use citation analysis to determine who are the leading voices in the field, with a view toward expanding the conversation.⁴ Alternatively, N-Gram analysis could be applied to digitized commentaries to

⁴ For an example of citation analysis to understand a genre of literature, in this case English literary studies, see David S. Nolen and Hillary A.H. Richardson, "The Search for Landmark

understand the shifting lexicon of Biblical studies, comparing that lexicon with other academic disciplines.⁵ The point is that there are a number of approaches that scholars and librarians could introduce together to show new ways of understanding the history of Biblical interpretation that do not fall victim to the canonization problem.

Jockers argues that, “Close reading is not only impractical as a means of evidence gathering in the digital library, but big data render it totally inappropriate as a method of studying literary history.”⁶ I would agree with Jockers’ sentiment; close reading depends too much on representative samples. Biblical scholars pride themselves on being careful, close exegetes, for good reason. It is important to ask, though, which texts they are choosing to perform this close reading on. Biblical studies, however, does not yet have the “big data” that Jockers refers to. Or, perhaps more accurately stated, Biblical studies has plenty of data, but it is not in the proper form for the type of reading that Jockers outlines. It is up to librarians to work with scholars and publishers not only to introduce these new forms of reading, but also to curate the data.

Works in English Literary Studies: A Citation Analysis,” *Journal of Academic Librarianship* 42.4 (2016): 453–58.

⁵ For an introduction to N-gram analysis in academic, see Burcu Caglar Gencosman, Huseyin C. Ozmutlu, and Seda Ozmutlu, “Character N-Gram Application for Automatic New Topic Identification,” *Information Processing & Management* 50.6 (2014): 821–56.

⁶ Jockers, *Macroanalysis*, 7.

REFERENCES

- Abruzzi, Ray. "Undertaking a Digitization Project?" *Online Searcher* 39, no. 2 (April 3, 2015): 28–32.
- Alexander, Marc, Fraser Dallachy, Scott Piao, Alistair Baron, and Paul Rayson. "Metaphor, Popular Science, and Semantic Tagging: Distant Reading with the Historical Thesaurus of English." *Digital Scholarship in the Humanities* 30 (December 2, 2015): i16–27. doi:10.1093/llc/fqv045.
- Best, Stephen, and Sharon Marcus. "Surface Reading: An Introduction." *Representations* 108, no. 1 (2009): 1–21. doi:10.1525/rep.2009.108.1.1.
- Blakesley, Jacob. "Examining Modern European Poet-Translators 'Distantly.'" *Translation & Literature* 25, no. 1 (March 2016): 10–27. doi:10.3366/tal.2016.0235.
- Breeding, Marshall. "Library Systems Report 2017." *American Libraries* 48, no. 5 (May 2017): 22–35.
- David J. A. Clines. "Writing a Job Commentary." In *The Genre of Biblical Commentary*, 29–39. Eugene, OR: Wipf and Stock, 2015.
- Dobson, James E. "Can an Algorithm Be Disturbed? Machine Learning, Intrinsic Criticism, and the Digital Humanities." *College Literature* 42, no. 4 (Fall 2015): 543–64.
- Goldstone, Andrew, and Ted Underwood. "The Quiet Transformations of Literary Studies: What Thirteen Thousand Scholars Could Tell Us." *New Literary History* 45, no. 3 (2014): 359–84.
- Gooding, Paul, Melissa Terras, and Claire Warwick. "The Myth of the New: Mass Digitization, Distant Reading, and the Future of the Book." *Literary & Linguistic Computing* 28, no. 4 (December 2013): 629–39.
- Greco, Gina L., and Peter Shoemaker. "Intertextuality and Large Corpora: A Medievalist Approach." *Computers and the Humanities* 27, no. 5/6 (1993): 349–55.

- Halsell, Lindsey A. "TEI and So Can You: Corpus Linguistics and the Text Encoding Initiative." *PMLA Quarterly* 77, no. 3 (Spring 2013): 63–70.
- Hanneken, Todd Russell. "Digital Archaeology's New Frontiers." *Biblical Archaeology Review* 43, no. 2 (March 2017): 28.
- Hardesty, Juliet L. "Transitioning from XML to RDF: Considerations for an Effective Move Towards Linked Data and the Semantic Web." *Information Technology & Libraries* 35, no. 1 (March 2016): 51–64. doi:10.6017/ital.v35i1.9182.
- Hayles, N. Katherine. "Combining Close and Distant Reading: Jonathan Safran Foer's Tree of Codes and the Aesthetic of Bookishness." *PMLA: Publications of the Modern Language Association of America* 128, no. 1 (January 2013): 226–31.
- . *How We Think: Digital Media and Contemporary Technogenesis*. Chicago ; London: University Of Chicago Press, 2012.
- Heuser, Ryan, and Long Le-Khac. "Learning to Read Data: Bringing out the Humanistic in the Digital Humanities." *Victorian Studies* 54, no. 1 (2011): 79–86. doi:10.2979/victorianstudies.54.1.79.
- Jockers, Matthew L. *Macroanalysis: Digital Methods and Literary History*. 1st Edition. Urbana: University of Illinois Press, 2013.
- Kirschenbaum, Matthew G. "The Remaking of Reading: Data Mining and the Digital Humanities." *Conference Proceedings; National Science Foundation Symposium on Next Generation of Data Mining and Cyber-Enabled Discovery for Innovation, Baltimore, October 10-12, 2007*. <https://www.csee.umbc.edu/~hillol/NGDM07/abstracts/talks/MKirschenbaum.pdf>.
- Lenski, R.C.H. *Interpretation of the Acts of the Apostles*. Columbus, O.: Wartburg Press, 1944.
- Liddle, Dallas. "Reflections on 20,000 Victorian Newspapers: 'Distant Reading' The Times Using The Times Digital Archive." *Journal of Victorian Culture (Routledge)* 17, no. 2 (June 2012): 230–37. doi:10.1080/13555502.2012.683151.
- . "Reflections on 20,000 Victorian Newspapers: 'Distant Reading' The Times Using The Times Digital Archive." *Journal of Victorian Culture (Routledge)* 17, no. 2 (June 2012): 230–37. doi:10.1080/13555502.2012.683151.
- Liu, Alan. "The Meaning of the Digital Humanities." *PMLA* 128, no. 2 (March 1, 2013): 409–23. doi:10.1632/pmla.2013.128.2.409.
- Miller, Larisa K. "All Text Considered: A Perspective on Mass Digitizing and Archival Processing." *The American Archivist* 76, no. 2 (2013): 521–41.

- Moretti, Franco. "Conjectures on World Literature." *New Left Review*, II, no. 1 (February 2000): 54–68.
- . *Distant Reading*. London ; New York: Verso, 2013.
- . *Graphs, Maps, Trees: Abstract Models for a Literary History*. London ; New York: Verso, 2005.
- . "Style, Inc. Reflections on Seven Thousand Titles (British Novels, 1740–1850)." *Critical Inquiry* 36, no. 1 (2009): 134–58. doi:10.1086/606125.
- Moretti, Franco, and Alberto Piazza. *Graphs, Maps, Trees: Abstract Models for Literary History*. London; New York: Verso, 2007.
- Morgan, Eric Lease. "An Introduction to the Search/Retrieve URL Service (SRU)." *Ariadne*, no. 40 (2004). <http://www.ariadne.ac.uk/issue40/morgan>.
- Mueller, Derek. "Grasping Rhetoric and Composition by Its Long Tail: What Graphs Can Tell Us about the Field's Changing Shape." *College Composition and Communication* 64, no. 1 (2012): 195–223.
- Muresan, Gheorghe, and David J. Harper. "Topic Modeling for Mediated Access to Very Large Document Collections." *Journal of the American Society for Information Science & Technology* 55, no. 10 (August 2004): 892–910. doi:10.1002/asi.20034.
- Ramsay, Stephen. *Reading Machines: Toward an Algorithmic Criticism*. 1st Edition edition. University of Illinois Press, 2011.
- Rankin, Sharon, and Casey Lees. "The McGill Library Chapbook Project: A Case Study in TEI Encoding." *OCLC Systems & Services* 31, no. 3 (July 2015): 134–43. doi:10.1108/OCLC-07-2014-0030.
- Ryan, Simon. "The Location of Knowledge." *International Journal of Technology, Knowledge & Society* 8, no. 6 (November 2012): 73–81.
- Sag, Matthew. "Orphan Works as Grist for the Data Mill Symposium: Orphan Works & Mass Digitization: Obstacles & Opportunities." *Berkeley Technology Law Journal* 27 (2012): 1503–50.
- Salminen, Airi, Eliisa Jauhiainen, and Reija Nurmeksela. "A Life Cycle Model of XML Documents." *Journal of the Association for Information Science & Technology* 65, no. 12 (December 2014): 2564–80. doi:10.1002/asi.23148.
- Skaggs, Bradley, and Lise Getoor. "Topic Modeling for Wikipedia Link Disambiguation." *ACM Transactions on Information Systems* 32, no. 3 (June 2014): 10–10:24. doi:10.1145/2633044.

- Tangherlini, Timothy R., and Peter Leonard. "Trawling in the Sea of the Great Unread: Sub-Corpus Topic Modeling and Humanities Research." *Poetics* 41, no. 6 (December 2013): 725–49. doi:10.1016/j.poetic.2013.08.002.
- Tanoukhi, Nirvana. "The Scale of World Literature." *New Literary History* 39, no. 3 (2008): 599–617.
- Taylor Young, Daniel. "How Do You Measure a Constitutional Moment? Using Algorithmic Topic Modeling To Evaluate Bruce Ackerman's Theory of Constitutional Change." *Yale Law Journal* 122, no. 7 (May 2013): 1990–2054.
- Trumpener, Katie. "Critical ResponseI. Paratext and Genre System: A Response to Franco Moretti." *Critical Inquiry* 36, no. 1 (2009): 159–71. doi:10.1086/606126.
- Unsworth, John. "What Is Humanities Computing and What Is Not?" *Jahrbuch Für Computerphilologie* 4 (n.d.): 71–83.
- Widner, Michael. "Text-Mining the Middle Ages | Manuscripts and Machines," January 25, 2016. <https://people.stanford.edu/widner/content/text-mining-middle-ages>.
- Yarchin, William. "Introduction to the Volume." In *The Genre of Biblical Commentary*, 1–25. Eugene, OR: Wipf and Stock, 2015.
- . *The Genre of Biblical Commentary: Essays in Honor of John E. Hartley on the Occasion of His 75th Birthday*. Edited by Timothy D. Finlay. Eugene, Oregon: Pickwick Publications, 2015.
- Zhou, Yongli. "Fulfill Your Digital Preservation Goals with a Budget Studio." *Information Technology & Libraries* 35, no. 1 (March 2016): 26–50. doi:10.6017/ital.v35i1.5704.
- Zukowski, Adam. "XML for Catalogers and Metadata Librarians." *Library Quarterly* 85, no. 3 (July 2015): 331–33.

APPENDICES

Appendix 1. Average Wordcount Data for 8 Commentaries on James

Scripture Reference	Average Number of Words	Standard Deviation of Number of Words
1:1	1900.125	1815.11774
1:2	949.125	939.8979333
1:3	717.625	682.868098
1:4	507.75	330.3784626
1:5	843.625	534.6916167
1:6	530.125	418.5077018
1:7	241.375	411.9004517
1:8	583	435.2326143
1:9	616	701.8213041
1:10	592	505.2886021
1:11	605.25	415.5769655
1:12	891.5	744.2140629
1:13	1025.5	808.0270505
1:14	420.375	389.9714046
1:15	419.75	329.8011306
1:16	408.625	556.9760029
1:17	1081.625	1175.220944
1:18	1199.875	1916.895588
1:19	1165.125	1136.720274
1:20	578.125	578.125
1:21	1034	1034
1:22	645.5	645.5
1:23	343.375	343.375
1:24	271.375	271.375
1:25	1059	1059
1:26	532	532
1:27	889.375	889.375

2:1	723.375	565.5566
2:2	336.375	331.5517359
2:3	390.25	296.0640616
2:4	548.625	446.0714028
2:5	797.25	607.8048912
2:6	452.625	291.8541404
2:7	332.875	203.7494522
2:8	556	381.41054
2:9	254.75	157.7980174
2:10	553.75	479.0770144
2:11	339.875	219.8333378
2:12	516.875	645.6094435
2:13	601.625	506.8919123
2:14	1018.75	819.7767553
2:15	170.875	199.9588797
2:16	321	262.5582928
2:17	265.5	133.194809
2:18	1135.375	987.5010217
2:19	520.875	356.9487508
2:20	333.875	242.4814794
2:21	1037	969.760206
2:22	455.125	416.9187785
2:23	683.125	648.4049523
2:24	723.125	630.4056376
2:25	585.75	403.6426107
2:26	316.5714286	165.0503531
3:1	695.375	522.3769677
3:2	844	1073.361742
3:3	313	278.826623
3:4	450.25	219.4947119
3:5	449.5	343.33865
3:6	1282.5	926.6845665
3:7	167.375	184.8211008
3:8	552.25	493.038031
3:9	530	504.5895078
3:10	238.625	155.6625815
3:11	185.875	183.2378918
3:12	417.625	513.51198
3:13	556.25	323.3064667

3:14	586.125	366.5789478
3:15	542	505.3400552
3:16	292.25	203.8751369
3:17	1084.125	1389.757783
3:18	474	408.8723168
4:1	710.75	426.1896794
4:2	1008	810.4090678
4:3	455.375	426.6744954
4:4	590.125	384.0656333
4:5	917.25	662.8056707
4:6	833.625	1185.525554
4:7	969.25	1503.115313
4:8	779.125	1118.576064
4:9	543.625	566.8884842
4:10	381.875	402.1103038
4:11	763.5	596.8261293
4:12	521.875	464.0401576
4:13	603.25	544.2162779
4:14	473.875	359.0998419
4:15	512	444.2197011
4:16	353.875	244.1483375
4:17	465.4285714	336.3856206
5:1	724.25	766.4695968
5:2	244.25	270.0432505
5:3	839.25	830.6088559
5:4	580.25	428.4566156
5:5	633.25	455.4475821
5:6	639.5	462.1901279
5:7	971.875	652.332614
5:8	370.25	347.6368179
5:9	443.75	353.3811176
5:10	464.125	354.7000493
5:11	1081.375	1053.687937
5:12	1086.125	552.7542563
5:13	642.375	509.4430958
5:14	1647.75	1153.61107
5:15	694.75	565.9393582
5:16	1192.125	1345.424462
5:17	453.25	938.3655623

5:18	574.75	287.9825144
5:19	340.875	341.0175687
5:20	1155.625	1496.408934

Appendix 2. Bibliography of the 8 Commentaries on James

Calvin, Jean. *Calvin's Commentary on the Epistle of James Newly Translated from the Original Latin. With Notes, Practical, Historical, and Critical.* Aberdeen: Printed by JChalmers and Co; and sold by ABrown, 1797.

Davids, Peter H. *The Epistle of James: A Commentary on the Greek Text.* New International Greek Testament Commentary. Grand Rapids, Mich.: Eerdmans, 1982.

Henry, Matthew. *An Exposition on the Old and New Testament In Five Volumes. ... Wherein Each Chapter Is Summed up in Its Contents; the Sacred Text Inserted at Large in Distinct Paragraphs; Each Paragraph Reduced to Its Proper Heads; the Sense Given, and Largely Illustrated: With Practical Remarks and Observations.* By Matthew Henry, Late Minister of the Gospel. Fourth Edition. London: Printed for John and Paul Knapton, 1737.

Manton, Thomas. *A Practical Commentary, or an Exposition with Notes on the Epistle of James.* 2nd edition. London: Printed by John Macock for Luke Fawne, 1653.

Martin, Ralph P. *James.* Word Biblical Commentary ; v. 48. Waco, Tex.: Word Books, 1988.

Moo, Douglas J. *The Letter of James: An Introduction and Commentary.* Pillar New Testament Commentary. Grand Rapids, Mich.: WBEerdmans, 2000.

Reicke, Bo Ivar. *The Anchor Bible: The Epistles of James, Peter, and Jude.* Garden City, N.Y: Doubleday, 1964.

Ropes, James Hardy. *A Critical and Exegetical Commentary on the Epistle of St. James.* International Critical Commentary on the Holy Scriptures of the Old and New Testaments ; 41. New York: Charles Scribner's Sons, 1916.

Appendix 3. An Example of an XML-Encoded Commentary

```
<?xml version="1.0"?>
<data>
  <chapter name="1">
    <verse name="1">
```

EPISTOLARY INTRODUCTION 1:1

THE greeting formula presents James the Just, the brother of the Lord, writing to Jewish Christian congregations scattered outside of his “home” district of Palestine. The author (whether James or an editor of material stemming from him; cf. Introduction, 12–13) uses a typical Greek greeting style, a form which appears in both literary (e.g. 1 Macc. 10:25; Euseb. Praep. Ev. 9.33–34) and actual (e.g. 1 Thes. 1:1) letters of the period.

(1) The author names himself James or Jacob (Hebrew *y’qb*). As has been argued above (Introduction, 5–6), such a title could refer to only one person in the early church—James the Just (contra Meyer and Kürzdorfer). This title itself is very interesting. The term *δοῦλοῖς* by no means unusual on the lips of an apostle (e.g. Rom. 1:1; Gal. 1:10; Phil. 1:1), coming as it does from Jewish literature (e.g. Gn. 32:10; Jdg. 2:8; Ps. 89:3 [88:4]; Is. 41:8; Je. 26:7; Am. 3:7). It is both an indication of humility, for the servant does not come in his own name, and of office, for the bearer of the title is in the service of a great king (cf. Mussner, 60–61). But the reference to *Ἰησοῦ Χριστοῦ* is unusual, for “Christ” is now not a title, but a name. That use of the word would be surprising for a Jew who spent most of his time in Jerusalem where the titular use would have been meaningful, but it is very similar to the normal Pauline (i.e. Hellenistic) usage. Thus it is an indication that this verse stems from a Hellenistic Christian, that the editor has at least heavily shaped it. There is no reason to suppose (contra Spitta) that *Ἰησοῦ Χριστοῦ* is an interpolation, for the form and order of words is also found elsewhere (e.g. Rom. 1:7; 5:1, 11; 1 Cor. 1:3; 2 Cor. 1:2; Gal. 1:3).

In using the phrase *αἱ δώδεκα φυλαί*, the author looks on the recipients of the epistle as the true Israel. The church has quite naturally appropriated the title, for it was the work of the Messiah to reestablish the twelve tribes (Je. 3:18; Ezk. 37:19–24; Pss. Sol. 17:28), and Christians recognized themselves as the true heirs of the Jewish faith (Romans 4; 1 Cor. 10:18; Gal. 4:21–31; Phil. 3:3). While such a term for the church would fall quite naturally from the lips of a Jewish Christian, Paul also uses it in his letters to gentile churches.

[James, p. 64]

The second part of the title, however, produces some controversy; the term *διασπορά* was used by Jews to indicate that part of Judaism living outside of Palestine (cf. K. L. Schmidt, TDNT II, 99–101), but what would such a term mean to Christians? On the one hand, it is possible that they adopted the term and used it metaphorically to indicate their state as “strangers and pilgrims” upon the earth (cf. Heb. 11:13; 13:14; and 1 Pet. 1:1, 17; 2:11). Thus Dibelius sees the book addressed to “the true Israel, whose home is in heaven, but for whom the earth is only a foreign land.” This would be the most likely meaning of the term if the work were written outside of Palestine (and a possible meaning if it were written within that land). On the other

hand, if one assumes not only a Palestinian, but a Jewish Christian provenance for the work, it would be unnatural to leap over the literal meaning. What other term would such a group have used to refer to Christians outside of Palestine, i.e. to Christians living “in the Diaspora”? This was the logical and natural way to do so. As a result, we conclude with Mussner that although the metaphorical sense is attractive, accepting as we do the Jewish Christian origin of this epistle, the most natural way of reading this phrase is as an address to the true Israel (i.e. Jewish Christians) outside of Palestine (i.e. probably in Syria and Asia Minor).

The greeting itself is at first glance not very significant, for it is the standard Greek epistolary greeting. But when compared with the Pauline greetings it is surprising. Paul uses the double formula *χάρις ὑμῖν καὶ εἰρήνῃ* which shows both Jewish-oriental (*šālôm*; cf. Dan. 4:1 Theod.) and Christian liturgical practice (in the style and use of *χάρι* cf. H. Conzelmann, TDNT IX, 393–394). Why are these influences absent from a Jewish Christian letter? The author has simply used the conventional Greek term, either because he lacked Paul’s creativity and mastery of Greek or because the Hellenistic redactor/scribe had principal responsibility for v 1 and did not think in Paul’s more Aramaic terms.

[James, p. 65]

</verse>