

A HIDDEN MARKOV MODEL APPROACH TO AUTHORSHIP ATTRIBUTION OF THE
PASTORAL EPISTLES

Josiah Leinbach

A Thesis

Submitted to the Graduate College of Bowling Green
State University in partial fulfillment of
the requirements for the degree of

MASTER OF SCIENCE

December 2024

Committee:

Shuchismita Sarkar, Committee Chair

Riddhi Ghosh

Christopher Rump

This content is unable to be made fully accessible
because it would result in a fundamental alteration of
the document.

Copyright © December 2024

Josiah Leinbach

All rights reserved

ABSTRACT

Shuchismita Sarkar, Committee Chair

The New Testament contains thirteen epistles written in the name of the Apostle Paul, and from the earliest records of church history, Christian theologians received all thirteen as authentically Pauline. Since the 19th century, however, many scholars have doubted Paul's authorship of some epistles based on, among other factors, their vocabulary and writing style, which differ from undisputed Pauline epistles. In particular, three epistles called the Pastoral Epistles (1 Timothy, 2 Timothy, and Titus) have been subject to the most doubt. This thesis will use a Hidden Markov Model that analyzes the transitions between different parts of speech in the whole Pauline corpus and classifies sentences as belonging to a "Pauline" or "non-Pauline" style. Then, informed by New Testament scholarship, we will interpret these results and judge the possibility of Pauline authorship for the Pastoral Epistles.

ACKNOWLEDGMENTS

A thesis such as this involves the sundry efforts of those who labored many months, and in some cases many years, to bring it to fruition.

I express great gratitude to my advisor, Dr. Sarkar, for her wonderful guidance, insights, and encouragement. Neither this thesis, nor my statistical acumen, would be even half of what they are without her work. The faculty of Hillsdale College, by their instruction and friendship, have likewise played no small part in my intellectual formation. For them as well, I give the utmost thanks.

I also thank my parents, who from my childhood poured themselves into my education, taught me to love learning, and inspired me to pursue excellence in all I did.

Indirectly, I am grateful for the Christian and Missionary Alliance's youth Bible Quizzing program, through which my younger self memorized more than half of the books analyzed in this thesis.

Finally, I do give most humble and hearty thanks to Almighty God, whom I thank for these and all other blessings He has given me in this life.

Soli Deo Gloria.

TABLE OF CONTENTS

	Page
Chapter 1 INTRODUCTION	1
1.1 Historical and Textual Background to the Pauline Corpus	1
1.2 Deutero-Pauline Epistles	2
1.3 Pastoral Epistles	3
Chapter 2 TESTING THEORY	6
2.1 Establishing a Functional Null Hypothesis	6
2.2 Structure of Tests and Interpretation of Results	8
Chapter 3 METHODOLOGY	11
3.1 Canon-Construction Methods	11
3.2 Criteria for Style Classification	12
3.3 Criteria for Style Significance	14
3.4 Statistical Methods	14
3.4.1 Hidden Markov Model (HMM)	14
3.4.2 Hidden Markov Model for Authorship Attribution	15
3.4.3 Emission Distribution for Authorship Attribution	16
3.4.4 Estimation of Model Parameters	16
Chapter 4 RESULTS	19
4.1 Hyper-Parameter Selection	19
4.1.1 Nomenclature	19
4.2 Style Classifications	20
4.2.1 $S_1=1, S_2=1$ — 13 Book Claimed Pauline Corpus	20
4.2.2 $S_1=1, S_2=1$ — 13 Book Claimed Pauline Corpus + Hebrews and 1 John	22
4.2.3 $S_1=4, S_2=2$	26
Chapter 5 INTERPRETATIONS	34
5.1 PS_4 and Subject Matter-Dependency	34

5.2 Application to the Pastoral Epistles	39
Chapter 6 CONCLUSIONS	43
Chapter 7 Bibliography	45

LIST OF FIGURES

Figure		Page
4.1	Solution A* ($S1=1, S2=1$)	21
4.2	Solution B* ($S1=1, S2=1$)	21
4.3	Solution A for ($S1=1, S2=1$) with Hebrews and 1 John	22
4.4	Solution B for ($S1=1, S2=1$) with Hebrews and 1 John	23
4.5	Solution C ($S1=4, S2=2$)	27
4.6	Solution C with ambiguities ($S1=4, S2=2$)	28
4.7	Frobenius Distance for Solution C ($S1=4, S2=2$)	32

LIST OF TABLES

Table	Page
3.1 POS Table	13
4.1 POS transition probabilities for Solution A (S1=1, S2=1)	24
4.2 POS transition probabilities for Solution B (S1=1, S2=1)	26
4.3 Comparison between transition matrices: PS1 vs. NPS1	26
4.4 POS transition probabilities for Solution B (S1=4, S2=2)	30
4.5 Transition probability differences greater than 0.1	33
5.1 Verse-breaks for personal greetings	35
5.2 Summary data on sentence length by style classification	36
5.3 Posterior probabilities for Ephesians	37
5.4 Posterior probabilities for Colossians	38
5.5 Posterior probabilities for 2 Thessalonians	38
5.6 Verse breaks for style transitions, based on posterior probabilities	39
5.7 Proportion for all POS in each style	41

PREFACE

“The superiority in judgement and diligence which you are going to attribute to the Biblical critics will have to be almost superhuman if it is to offset the fact that they are everywhere faced with customs, language, race-characteristics, class-characteristics, a religious background, habits of composition, and basic assumptions, which no scholarship will ever enable any man now alive to know as surely and intimately and instinctively as the reviewer can know mine. And for the very same reason, remember, the Biblical critics, whatever reconstructions they devise, can never be crudely proved wrong. St. Mark is dead. When they meet St. Peter, there will be more pressing matters to discuss.”

— C. S. Lewis, *Modern Theology and Biblical Criticism* (1959)

CHAPTER 1 INTRODUCTION

1.1 Historical and Textual Background to the Pauline Corpus

Of the twenty-seven books in the New Testament (NT) canon, thirteen claim as their author the Apostle Paul, or Saint Paul, a 1st century Pharisee and one-time persecutor of the Christian church, who then converted to Christianity. These books, called the Pauline epistles, are, in canonical order: Romans, 1 Corinthians, 2 Corinthians, Galatians, Ephesians, Philippians, Colossians, 1 Thessalonians, 2 Thessalonians, 1 Timothy, 2 Timothy, Titus, and Philemon. Their canonical order differs from their (likely) chronological order on account of the canonical order dividing the books into ecclesiastical epistles written to churches (the first nine), and personal epistles written to individuals (the last four). Within each group, books were ordered from longest to shortest. Dates for their authorship are at times imprecise and vary among scholars, but a reasonable chronological order would be: Galatians, 1 Thessalonians, 2 Thessalonians, 1 Corinthians, 2 Corinthians, Romans, Philippians, Philemon, Colossians, Ephesians, Titus, 1 Timothy, 2 Timothy.

The first several generations of post-apostolic theologians, known as the Ante-Nicene Fathers—some of whom (e.g. Polycarp) were disciples of the Apostles themselves—universally attested to Pauline authorship of all thirteen epistles, oftentimes attributing the anonymous Epistle to the Hebrews to Paul as well. Subsequent church fathers from both the Eastern and the Western church held the same position. Within the thirteen book Pauline corpus, one finds a wide Greek vocabulary and multiple literary styles across different letters, something not lost on the church fathers. Yet where they attributed these different styles to the work of a single, versatile author in Saint Paul, many scholars from the early 19th century onward have believed them the product of multiple authors, of which Paul was one. How many of the thirteen books scholars believe Paul wrote varies, with numbers that range from just four to all thirteen, though at present a widespread consensus holds that he wrote at least seven: Romans, 1 Corinthians, 2 Corinthians, Galatians, Philippians, 1 Thessalonians, and Philemon. These “Undisputed Seven” stand against the three “Deutero-Pauline Epistles” (Ephesians, Colossians, and 2 Thessalonians), and also the

Pastoral Epistles (1 Timothy, 2 Timothy, and Titus).

1.2 Deutero-Pauline Epistles

Though the 16th century humanist scholar Erasmus of Rotterdam raised questions about Pauline authorship of Ephesians, not until the early 19th century did scholars begin to doubt Paul's authorship of Ephesians, and with it Colossians. Ephesians and Colossians share similar themes and vocabulary that are either absent in other claimed Pauline epistles, or else not discussed in such depth or from such perspectives. But the deciding difficulty with Ephesians and Colossians comes from their sentence structure, which are far longer and more syntactically complex than most books in the New Testament. After the introductory greeting, Ephesians' first sentence continues on for almost 260 words, something that, while more commonly done in Greek than in English, stands out as unique in the Pauline corpus. Ephesians averages about twice the number of words per sentence than Galatians, and Colossians is similarly more verbose. 2 Thessalonians does not display such divergent syntax, yet it still deviates somewhat, and its themes have been described as contradictory to what is found in the Undisputed Seven. For example, some scholars have argued that the eschatology presented in 2 Thessalonians 2 teaches that Christ will return at some point in the distant future, preceded by various signs; whereas 1 Thessalonians teaches that Christ's return is imminent, and that Paul himself expects to be alive at that point.

Yet despite these contrasts in literary character, the majority of scholars still hold to Pauline authorship of Ephesians, Colossians, and 2 Thessalonians. Ephesians can be seen as working out very Pauline themes of justification and atonement, just from different vantage points. In his exegetical commentary on Ephesians, Dr. Harold Hoehner surveyed 279 other scholarly commentaries from the last two centuries and showed not only that support for Pauline authorship never fell below forty-five percent in a given decade, but also that majority support for Pauline authorship has been the norm.¹ In fact, many notable scholars still today defend the authenticity of all three epistles.² Journals publish scholarly papers that either argue for, or at least

¹Harold Hoehner, *Ephesians: An Exegetical Commentary* (Baker Publishing Group, 2001), 20.

²For examples of recent commentaries defending Pauline authorship of Deutero-Pauline epistles, see S. M. Baugh, *Ephesians: Evangelical Exegetical Commentary* (Lexam Press, 2016); G. K. Beale, *Colossians and Philemon: A*

assume, Pauline authorship of 2 Thessalonians,³ and close-readings of the text and consistent exegesis have resolved its apparent contradictions with 1 Thessalonians.⁴

1.3 Pastoral Epistles

The Pastoral Epistles (PE) pose a separate problem. In terms of content, they contain themes of church governmental hierarchy, administrative matters, and the need to preserve sound doctrine; themes that many scholars believe evince a more developed tradition than could have existed in the early to mid A.D. 60s., when Paul is supposed to have written the PE. They also have no conspicuous placement in the chronology of Paul's missionary journeys related in the Acts of the Apostles, a historical narrative written by Paul's fellow missionary and chronicler Luke (sometimes called Saint Luke the Evangelist). For those reasons, many scholars reject the PE as authentically Pauline. This evidence is called "external evidence" because it relies on historical arguments outside the text in question. Complementing external evidence is "internal evidence", which mostly considers matters of grammar and style within the text.⁵

The relationship between external and internal evidence, however, is not always straightforward, complicating debates over authorship. With respect to external evidence of the PE, scholarship has demonstrated that the church in Paul's lifetime was more developed, both in governmental structure and doctrine, than some have previously indicated.⁶ Furthermore, the universal attestation of Pauline authorship of the PE by the church fathers is itself compelling, external evidence *in support* of Pauline authorship, as these men lived close to the books' time of writing, sometimes being only one degree removed from Paul himself (e.g. Polycarp was a

Paragraph-by-Paragraph Exegetical Evangelical Bible Commentary (Baker Academic Press, 2019); Gordon Fee, *The First and Second Letters to the Thessalonians* (*New International Commentary on the New Testament*) (Eerdmans, 2009). Elsewhere in their non-monographic works, these scholars have argued for Pauline authorship of the other Deutero-Pauline epistles.

³Paul Foster, "Who Wrote 2 Thessalonians? A Fresh Look at an Old Problem", *Journal for the Study of the New Testament* 35, no. 2 (2012). <https://doi.org/10.1177/0142064X12462654>.

⁴Sydney E. Tooth, "Suddenness and Signs: The Eschatologies of 1 and 2 Thessalonians" (PhD diss., University of Edinburgh, 2019).

⁵Internal evidence also considers internal cohesion, and consistency (textually and biographically) across the individual work and the corpus; but the methodology by which to approach those matters is more in keeping with the methodology of external evidence.

⁶George W. Knight III, *The Pastoral Epistles: A Commentary On the Greek Text* (Eerdmans, 1992), 35-36; Philip H. Towner, *The Letters to Timothy and Titus* (Eerdmans, 2006), 50-52.

disciple of the Apostle John). They also were native Koine Greek speakers better attuned to the language and literary details that scholars today must reconstruct artificially.

A “forgery” or “pseudonym” hypothesis, while perhaps an attractive solution, has serious theoretical and empirical difficulties. First, the early church detected and condemned many forged or pseudonymous documents written in the name of the apostles, so it is unlikely that the PE would have been accepted as Pauline if there were doubts as to whether Paul, in fact, wrote them.⁷ Second, of those detected and denounced forgeries, none had been accepted as canonical works.⁸ Therefore, if one assumes the works are not authentic to Paul, then the only plausible theory is that the early church was incorrect in their assessment and mistook them as being Pauline, itself an unlikely proposition. In sum, the external evidence and the assumptions underlying it incline in favor of Pauline authorship.

Yet the internal evidence seems to stand in opposition to the external evidence. The vocabulary of the PE differs from that of the Undisputed Seven, and contains proportionally more *hapax legomena*⁹ than the other Pauline works. Though some scholars have argued that the PE’s language is not that dissimilar from Paul’s other epistles, that so much attention is given to the linguistic differences underscores the differences that exist. Nor is PE’s literary form of a widely-known type from the period, which makes them difficult to assess. The New Testament scholar Luke Timothy Johnson has categorized 1 Timothy and Titus specifically as a type of *mandata principis*, or “commandments of the ruler”, letters that generally gave commands, instructions, and advice to their delegates or subordinates.¹⁰ As a form, the *mandata principis* is debated and somewhat nebulous on account of a dearth of extant examples, a fact that, regardless the merits of Johnson’s arguments, must qualify their strength.

⁷Michael J. Krueger, “The Authenticity of 2 Peter”, *Journal of Evangelical Theological Society* 42, no. 4 (1999), 648. “These letters [the Epistles to Laodicea and Alexandria], although claiming to be from the apostle Paul, were rejected as ‘forgeries’—hardly a term that suggests pseudonymous works were looked upon with acceptance by the early church. The document goes on to declare that these epistles ‘cannot be received into the catholic church, since it is not fitting that poison be mixed with honey.’”

⁸J.A.T. Robinson, *Redating the New Testament* (Wipf and Stock, 2000), 187.

⁹Unique words in a corpus.

¹⁰Luke Timothy Johnson, *The First and Second Letters to Timothy: A New Translation with Introduction and Commentary* (Yale University Press, 2008), 139-142; 159-160.

So while the claim of Pauline authorship for the PE seems probable under the external evidence, it seems improbable under the internal evidence, leaving the debate over authorship stalled.

CHAPTER 2 TESTING THEORY

In the midst of this standstill, this thesis will perform a stylostatistical analysis of the PE's Greek text, as a contribution to ongoing debates about their Pauline or non-Pauline authorship. Such an undertaking requires a testing theory consistent with the nature of the materials and the problem at hand.

2.1 Establishing a Functional Null Hypothesis

Although this thesis will conduct no formal hypothesis test, all investigations into Pauline authorship assume a *de facto* “null hypothesis” against which data are weighed and conclusions reached. Much of the impasse on the PE's authenticity comes from an epistemologically improper null hypothesis that takes for granted their *non-Pauline* authorship. Given that this assumption dominates popular and scholarly conceptions of the PE, something must be said on this before continuing.

Under the assumption of *non-Pauline* authorship, Pauline authorship must essentially be proved rather than disproved. But this inverts the matter and creates a dilemma. A compound null hypothesis of authorship increases the required evidence to reject the null hypothesis. Let us assume we have three possible defined authors (A , B , and C) being tested according to the logical proposition:

$$1. (B \vee C) \vee A$$

$(B \vee C)$ serves as a compound null hypothesis, and A a simple alternative hypothesis. The symbol $\neg$ stands for negation (“not”) and $\vee$ for a disjunctive (“or”). Next, we assume,

$$2. \neg B \text{ — (Assumed premise).}$$

Based on these two assumed premises, we can proceed through the syllogism and conclude:

$$3. B \vee (C \vee A) \text{ — (1: Associativity).}$$

$$4. C \vee A \text{ — (2, 3: Disjunctive Syllogism)}$$

5. $\neg C$ — (*Assumed premise*).

6. $\therefore A$ — (**4, 5: Disjunctive Syllogism**)

We reject the null hypothesis only if we have sufficient evidence to reject both B and C as possible authors. As a generalization, if the null hypothesis for authorship contains n possible authors, then each of the n authors must be tested.

The problem for the PE should now be evident, in that a compound null hypothesis without a defined author becomes substantially more challenging to test against, and only possible when the set of all possible authors can be defined and have known, extant works. Herein lies the second problem for those who assume a null hypothesis of non-Pauline authorship, as a theoretically infinite set of people, known and unknown, could have authored the PE. This set necessarily includes figures lost to history (e.g. an unknown disciple of Paul's in Ephesus), and known figures with no known extant writings (e.g. Apollos, mentioned in Acts, 1 Corinthians, and Titus). No sound test can be performed in that case, for there will always exist at least one conceivable, theoretical author, A_i , whose style more closely matches the PE than any known author, namely Paul. If any investigation into authorship of the PE begins by assuming non-Pauline authorship as its *de facto* or *de jure* null hypothesis, it cannot devise a test to reject the unknown number of unknown authors, and thus in theory cannot (and in practice does not) reject that null hypothesis. A null hypothesis so constructed is unfalsifiable, and accordingly cannot serve as a starting assumption.

Instead, investigations into Pauline authorship of the PE should proceed from a null hypothesis that affirms Pauline authorship, rather than rejects it. All three Pastoral Epistles begin by claiming Paul as the author (1 Tim. 1:1, 2 Tim. 1:1, Tit. 1:1) and never circulated under another name. The early church testified unanimously to Pauline authorship on the belief that he was the real author behind them. Finally, unlike the compound, unfalsifiable null hypothesis assumed most often today, a null hypothesis of Pauline authorship is a simple null hypothesis that can be falsified. For these reasons, this thesis will proceed with a stated null hypothesis that Paul wrote the PE. Bankers test for counterfeit bills by first assuming the bill is genuine, and then

administering tests that would mark it as fake. The same process should apply here as well. (It should be noted that operating from this null hypothesis makes no assumptions about its correctness. That shall be discussed in due course.)

2.2 Structure of Tests and Interpretation of Results

Reframing the null hypothesis this way shifts what will be searched for in the tests themselves. A null hypothesis that assumes non-Pauline authorship cannot be falsified, so in practice most investigations try to prove the alternative hypothesis instead; an endeavor that, besides the obvious methodological problems, forces scholars to look for stylistic similarities between the PE and other Pauline letters. Returning to the previous analogy of counterfeit bills, bankers are trained to detect counterfeits by studying genuine currency, since a counterfeit bill could take on any number of characteristics. This becomes all the more true with a complex task like authorship attribution, in which multiple known, and perhaps even unknown, genuine styles interact with one another. For the PE in particular, understanding the *differences* between them and the Undisputed Seven will be of greater value than understanding the similarities, for insights are more likely to be derived from anomalies than normality. Tests on the PE, therefore, will focus on finding differences between them and the Undisputed Seven.

But a mere tally of differences in vocabulary is insufficient to decide for or against Pauline authorship of the PE for three reasons. First, different styles do not necessarily imply different authors. This is the deficiency in P. N. Harrison's work *The Problem of the Pastoral Epistles* (1921),¹ in which Harrison argued that the PE's vocabulary is not sufficiently Pauline. More than a century later, Harrison's arguments still exert disproportionate influence on scholarly arguments against Pauline authorship of the PE, despite his study's faults and limitations.² On the subject of stylostatistical methods like those of Harrison, the famed 20th century New Testament scholar Bruce Metzger wrote that scholars must wrestle with several questions. With respect to

¹P.N. Harrison, *The Problem of the Pastoral Epistles* (Wipf & Stock, 2016).

²For a critique of Harrison's work, see Jermo van Nes, "The Problem of the Pastoral Epistles: An Important Hypothesis Reconsidered", ed. Stanley E. Porter and Gregory P. Fewster, *Paul and Pseudopigraphy* (Brill, 2013), 153-169.

vocabulary studies, Metzger asks, “How different can the results of a particular analysis of the two texts be before they throw serious doubt upon the theory that they have a common author?”³ Opponents and proponents of Pauline authorship have struggled to answer this question.

Second, various textual circumstances may influence the PE’s style in ways hard to discern. Metzger continues: “What allowance should be made, in assessing specific texts, for differences in the two works as regards (1) subject matter and (2) literary form? If the subject and form are different, can the investigator devise a set of tests which are least likely to be disturbed by this?”⁴ Disentangling subject and form from style is a delicate task, one that requires deep and integrated knowledge of statistical methods, New Testament scholarship, and the New Testament text itself.

Third, the PE were written two-thousand years ago. Insofar as our knowledge of ancient stylistic, rhetorical, and grammatical practices are concerned, there are many “unknown unknowns” that elude, if not render impossible, objective and *a priori* solutions.⁵ For that reason, authorship attribution of New Testament books cannot be treated as a mere statistics problem. Instead, it must be informed by New Testament scholarship that accounts for non-quantifiable variables like a book’s content, theology, literary form, etc. With that in mind, verdicts against Pauline authorship of a book must meet two necessary criteria, one quantitative and one qualitative:

1. **Quantitative:** A majority of its sentences are classified as non-Pauline.
2. **Qualitative:** All theories that assume Pauline authorship lack sufficient explanatory power to offer a plausible explanation for the results obtained in Criterion #1.

The latter of these two serves as a safeguard against reducing the question to a mere statistics problem that ignores New Testament scholarship. Even so, scholarship’s strength is not left unchecked, since a large enough majority of non-Pauline sentences will automatically and

³Bruce Metzger, “A Reconsideration of Certain Arguments Against the Pauline Authorship of the Pastoral Epistles”, *The Expository Times*, 70, no. 3 (December 1958): 93.

⁴Ibid.

⁵Metzger, 93-94.

dramatically reduce the explanatory power of almost any explanatory theory that espouses Pauline authorship.

Should one or both criteria fail to be met, then the null hypothesis of Pauline authorship will not be rejected.

CHAPTER 3 METHODOLOGY

In keeping with these principles, this thesis will employ a complementary methodology that relies on both New Testament scholarship and robust statistical analysis.

3.1 Canon-Construction Methods

Of first importance is the matter of canon construction and determining what ought to be done with the thirteen claimed books in the Pauline corpus, as this will dictate the nature of the testing procedure. (Canon comes from the Greek meaning "rule" or "standard".) There are two general approaches for how one can build authorship-attribution models using these books.

The first approach assumes no prior knowledge about which books Paul may or may not have written. This we may call the "Naive Canon" or "Neutral Canon" approach. Because the Naive Canon approach makes no assumptions about what books are Pauline or not Pauline, it searches for books that are similar to one another, with the goal of finding clusters that could be classified as Pauline vs. non-Pauline. Tests performed under a Naive Canon limit the influence that prior beliefs about Pauline authorship exert on the model, and so are, in theory, more likely to yield unbiased results.

Yet that rejection of prior beliefs about authorship limits the strength and extent of conclusions reached using a Naive Canon. As already discussed, test results must be interpreted in conjunction with New Testament scholarship, and applying that scholarship to interpret test results becomes more challenging if it did not inform the process which produced said results. Put another way, its use as a prior and posterior tool are not easily separated. A Naive Canon could produce results wholly inconsistent with any proposed scholarly theory of Pauline authorship. Suppose, for example, that the Naive Canon classified Galatians, an Undisputed Epistle, as being non-Pauline. In that case, either 2,000 years of reception history confirmed by 400 years of New Testament scholarship is wrong, or the model has missed something. Common sense would lead us to conclude that the latter, if only because Paul's very existence in history assumes he wrote Galatians and the copious autobiographical material it contains. The possibility of unintelligible

results means that any informative results could well have happened by chance, thereby decreasing their explanatory power.

The second approach, and the one employed in this thesis, assumes some canon of authentic Pauline books that serve as the baseline canon to which other books are judged for admission to the canon. This we will call an Informed Canon approach. Unlike the Naive Canon, the Informed Canon assumes minimal prior beliefs about authorship, which puts it at far less risk of producing inexplicable results. Consequently, this thesis assumes two premises.

1. First, the Undisputed Seven are a reasonable and defensible lowest-common-denominator for the Informed Canon approach.
2. Second, for the Pastoral Epistles, stylistic similarity to—and, where applicable, plausibly explained variation from—the Undisputed Seven serve as sufficient criteria for consideration as “Pauline”.

3.2 Criteria for Style Classification

Most statistical inquiries into the authorial style of New Testament books have been lexical studies, that is, vocabulary-based studies. Vocabulary studies, based on counts of specific words and specific kinds of words, have held sway due to their relative simplicity and easy interpretation, but they have three shortcomings. First, specific vocabulary depends in large part on form and subject matter, as words and even usage rates of words will vary across genres and content. Second, one-dimensional studies, of which vocabulary studies are a type, do not account for more complex linguistic features like word order. Third, there is presently little agreement about what the Pauline vocabulary dictionary should contain, especially as Paul is widely believed to have invented new Greek words himself (e.g. θεοδιδακτος), showing his facility as a writer.¹ Comparatively few morphological studies have been performed, and these have the greater potential to reveal information about writing styles than lexical studies.²

¹Stephen E. Whitmer, “θεοδιδακτος in 1 Thessalonians 4:9: A Pauline Neologism”, *New Testament Studies* 52, no. 2 (2006), 239. <https://doi.org/10.1017/S0028688506000142>.

²Kenneth J. Neumann, *The Authenticity of the Pauline Epistles in the Light of Stylostatistical Analysis*, (Scholars Press, 1990), 44-50.

Part-of-Speech Tags	Original Label	Combined Label
Adjective	A	AD
Conjunction	C	I
Adverb	D	AD
Interjection	I	I
Noun	N	N
Preposition	P	I
Definite article	RA	RA
Demonstrative pronoun	RD	P
Interrogative/indefinite pronoun	RI	P
Personal pronoun	RP	P
Relative pronoun	RR	P
Verb	V	V
Particle	X	I

Table 3.1 POS Table

The study performed here uses the parts-of-speech (POS) used in the Pauline epistles and additional books used as controls, based on a merged text created from Society of Biblical Literature's Greek New Testament (SBLGNT) and the SBL morphological text. The former text contains the New Testament books in normal sentence form, and the latter contains the grammatical parsing for each word in the sentence. Word order, and therefore POS order, is more fluid in Greek than in English, which means differences in sentence construction are evidence for potentially different writing styles. Those differences, when combined with knowledge of the underlying texts and New Testament scholarship, form the necessary framework for interpreting the results.

Greek is a highly inflected language with five grammatical cases (nominative, vocative, genitive, dative, and accusative) and grammatical genders (masculine, feminine, neuter). We will not consider the specific grammatical parsing but only the POS tags (see Table 3.1).

A model with all thirteen POS tags would likely suffer from over-parameterization and thus produce weak results. To avoid over-parameterization in the selected models, these thirteen POS tags have been combined into six based on their grammatical relationships. The four tags for different kinds of pronouns have been combined into one single pronoun tag *P*. Adjectives and adverbs, as descriptive modifiers, have been combined into *AD*. Finally, given their fluid use in

Greek syntax and that they are often undeclined, conjunctions, interjections, prepositions, and particles have been combined into *I*.

3.3 Criteria for Style Significance

No statistical model can account for all or even most linguistic features, especially for an inflected language like Greek. As such, we must address the objection that this model ignores other important elements of Greek writing style, like the number of relative clauses, verbal infinitives, etc. This we grant as true, but trivially so, since it applies to every model in some degree. Rather, the question that faces every model is two-fold: (1) Whether it is sufficient to capture at least *some* significant differences in writing style, and (2) whether those differences have meaningful explanations? A joint-affirmative answer fulfills the necessary requirements for classifying a distinct writing style, and as long as a model is internally consistent, its results are no less valid than any other model.

3.4 Statistical Methods

3.4.1 Hidden Markov Model (HMM)

In an HMM, an underlying hidden state sequence dictates the observable outputs, known as emissions. Each hidden state independently generates observable outputs based on emission probabilities. The transition between the hidden states follows a Markov process. Typically, a first-order Markov chain is used for modeling this transition (*i.e.* the present state is only dependent on the one preceding it). Let $\mathbf{S} = \{S_t\}_{t=1}^T$ denote a sequence of unobserved (hidden) random variables, each with a finite state space $\{1, \dots, J\}$, and let $\mathbf{O} = \{O_t\}_{t=1}^T$ denote a corresponding set of observed random vectors. A hidden Markov model has the functional form

$$P(\mathbf{S}, \mathbf{O}) = P(S_1) \prod_{t=2}^T P(S_t|S_{t-1}) \prod_{t=1}^T P(O_t|S_t), \quad (3.1)$$

where $P(S_1)$ is the initial state distribution, $P(S_t|S_{t-1})$ represents the transition from state S_{t-1} to state S_t , and $P(O_t|S_t)$ is known as the emission distribution.

3.4.2 Hidden Markov Model for Authorship Attribution

Suppose there are $d_1 + d_2$ documents under consideration. The first d_1 documents are confirmed to be written by the author, but for the next d_2 documents, the authorship authenticity is doubtful. Let $\mathbf{Y} = \{y_{ij}\}$ be the set of documents where y_{ij} denote the j^{th} sentence of the i^{th} document with $i = 1, 2, \dots, d_1 + d_2$ and $j = 1, 2, \dots, N_i$. Within each document, we assume that j^{th} sentence is only dependent on the previous sentence.

It is reasonable to assume that one author can write in multiple styles, so let there be S_1 styles attributed to the d_1 confirmed authorship documents. The documents with doubtful authorship may adopt one of these S_1 styles, or one from a new set of S_2 styles. Let $\mathbf{W} = \{W_{ij}\}$ denote the set of hidden states (styles) where W_{ij} is associated with the j^{th} sentence of the i^{th} document. It should be noted that the confirmed authorship documents are restricted to S_1 styles only i.e. $W_{ij} \in \{1, 2, \dots, S_1\}$ for $i = 1, 2, \dots, d_1$. However, the documents with doubtful authorship may relate to any styles i.e. $W_{ij} \in \{1, 2, \dots, S_1 + S_2\}$ when $i = d_1 + 1, \dots, d_1 + d_2$.

Consider a first order Markov chain with initial state distribution $\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_{S_1+S_2})$ and transition probability matrix $\mathbf{\Lambda}$ with elements:

$$\{\lambda_{rs}\}_{r=1, \dots, S_1+S_2; s=1, \dots, S_1+S_2}.$$

The related restrictions are $\sum_{r=1}^{S_1+S_2} \pi_r = 1$ and $\sum_{s=1}^{S_1+S_2} \lambda_{rs} = 1$. Let $f(\cdot; \theta_r)$ be the emission distribution associated with the r^{th} hidden state.

The complete data likelihood for the hidden Markov model can be written as

$$\begin{aligned} \mathcal{L}_C(\boldsymbol{\Theta}; \mathbf{Y}, \mathbf{W}) = & \left(\prod_{i=1}^{d_1} \prod_{r=1}^{S_1} (c^* \mathbf{B} \boldsymbol{\pi})_r^{I(W_{i1}=r)} \prod_{j=2}^{N_i} \prod_{r=1}^{S_1} \prod_{s=1}^{S_1} (c^{**} \mathbf{B} \mathbf{\Lambda} \mathbf{B})_{rs}^{I(W_{ij}=s, W_{i(j-1)}=r)} \prod_{j=1}^{N_i} \prod_{r=1}^{S_1} f(y_{ij}; \theta_r)^{I(W_{ij}=r)} \right) \\ & \times \left(\prod_{i=d_1+1}^{d_1+d_2} \prod_{r=1}^{S_1+S_2} \pi_r^{I(W_{i1}=r)} \prod_{j=2}^{N_i} \prod_{r=1}^{S_1+S_2} \prod_{s=1}^{S_1+S_2} \lambda_{rs}^{I(W_{ij}=s, W_{i(j-1)}=r)} \prod_{j=1}^{N_i} \prod_{r=1}^{S_1+S_2} f(y_{ij}; \theta_r)^{I(W_{ij}=r)} \right), \end{aligned} \quad (3.2)$$

where $I(\cdot)$ denotes the indicator function. $\mathbf{B} = \begin{pmatrix} \mathbf{I}_{S_1} & \mathbf{0}_{S_1 \times S_2} \\ \mathbf{0}_{S_2 \times S_1} & \mathbf{0}_{S_2 \times S_2} \end{pmatrix}$, where $\mathbf{I}_n$ is an identity matrix of size n and $\mathbf{0}_{n_1 \times n_2}$ is a matrix of zeros of respective size. c^* and c^{**} represent normalizing constant and vector, respectively. The first part of the likelihood function applies to the undisputed documents which can follow S_1 styles only. $\mathbf{B}$, c^* and c^{**} help to make required adjustment to $\boldsymbol{\pi}$ and $\boldsymbol{\Lambda}$ for these documents. The second part of the likelihood is associated with the disputed documents which are free to adopt any style from the available set of $S_1 + S_2$ varieties.³

3.4.3 Emission Distribution for Authorship Attribution

As discussed in section 3.2, using parts of speech tagging is a suitable approach for modeling the authorship style in this case. Let y_{ijl} denote the l^{th} word of sentence y_{ij} where $l = 1, 2, \dots, n_{ij}$. We will model the emission distribution of the r^{th} hidden state as a first order Markov chain with K states (parts of speech). Under this model, the l^{th} word depends on its previous word only. Let $\boldsymbol{\alpha}_r = (\alpha_{r1}, \alpha_{r2}, \dots, \alpha_{rK})$ represent the initial probability vector for the r^{th} hidden state. The transition probability matrix for the same is defined as $\boldsymbol{\Gamma}_r$ with elements $\{\gamma_{rkm}\}_{k,m \in 1,2,\dots,K}$. The emission distribution for the r^{th} hidden state is then given by

$$f(y_{ij}; \theta_r) = \left(\prod_{k=1}^K \alpha_{rk}^{I(y_{ij1}=k)} \right) \left(\prod_{k=1}^K \prod_{m=1}^K \gamma_{rkm}^{x_{ijkm}} \right), \quad (3.3)$$

where x_{ijkm} is the number of transitions from state k to state m for sentence y_{ij} .

3.4.4 Estimation of Model Parameters

This model's parameters are estimated using EM algorithm, an iterative algorithm consisting of *E*xpectation and *M*aximization steps.⁴ In the E-step, the expectation of the complete log-likelihood given $\mathbf{Y}$, known as Q-function is computed. For the HMM, this involves

³L.E. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains", *The Annals of Mathematical Statistics* 41, no. 1 (1970): 164–171.

⁴A.P. Dempster, N.M. Laird, and D.B. Rubin, "Maximum likelihood for incomplete data via the EM algorithm (with discussion)", *Journal of the Royal Statistical Society, Series B* 39, no. 1 (1977): 1–38.

computing the following posterior probabilities using the forward-backward algorithm:

- $\eta_{ijr} = P(W_{ij} = r | \mathbf{Y}_i)$ is the probability that the j^{th} sentence of the i^{th} document belongs to the r^{th} hidden state,
- $\zeta_{ijrs} = P(W_{ij} = s, W_{i(j-1)} = r | \mathbf{Y}_i)$ is the joint probability that for the i^{th} document, the $(j-1)^{th}$ sentence belongs to the r^{th} hidden state and the j^{th} sentence belongs to the s^{th} hidden state.⁵

In the M-step, the initial probability vectors and the transition probability matrices are estimated as follows (in these expressions, single dot and double dot above a parameter denote the estimates from the previous and current iterations):

The estimation of the elements of the initial probability vector and transition probability matrices for the r^{th} hidden state are given by

$$\ddot{\pi}_r = \begin{cases} \frac{\sum_{i=1}^{d_1+d_2} \dot{\eta}_{i1r}}{d_1 + d_2} & \text{if } 1 \leq r \leq S_1, \\ \frac{\sum_{i=d_1+1}^{d_1+d_2} \dot{\eta}_{i1r}}{d_2} & \text{if } S_1 \leq r \leq S_1 + S_2, \end{cases} \quad (3.4)$$

$$\ddot{\lambda}_{rs} = \begin{cases} \frac{\sum_{i=1}^{d_1+d_2} \sum_{j=2}^{N_i} \dot{\zeta}_{ijrs}}{\sum_{i=1}^{d_1+d_2} \sum_{j=1}^{N_i} \dot{\eta}_{ijr}} & \text{if } 1 \leq r \leq S_1 \text{ \& } 1 \leq s \leq S_1, \\ c^{***} \sum_{i=d_1+1}^{d_1+d_2} \sum_{j=2}^{N_i} \dot{\zeta}_{ijrs} & \text{otherwise} \end{cases} \quad (3.5)$$

where c^{***} is a normalizing constant.

The estimates of the parameters of the emission distribution are as follows:

$$\ddot{o}_{rk} = \begin{cases} \frac{\sum_{i=1}^{d_1+d_2} \sum_{j=1}^{N_i} \dot{\eta}_{ijr} I(y_{ij1}=k)}{\sum_{i=1}^{d_1+d_2} \sum_{j=1}^{N_i} \dot{\eta}_{ijr}} & \text{if } 1 \leq r \leq S_1, \\ \frac{\sum_{i=d_1+1}^{d_1+d_2} \sum_{j=1}^{N_i} \dot{\eta}_{ijr} I(y_{ij1}=k)}{\sum_{i=d_1+1}^{d_1+d_2} \sum_{j=1}^{N_i} \dot{\eta}_{ijr}} & \text{if } S_1 \leq r \leq S_1 + S_2, \end{cases} \quad (3.6)$$

⁵L.R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition", *Proceedings of the IEEE* 77, no. 2 (1989): 257–286.

$$\ddot{\gamma}_{rkm} = \begin{cases} \frac{\sum_{i=1}^{d_1+d_2} \sum_{j=1}^{N_i} \dot{\eta}_{ijr} x_{ijkm}}{\sum_{i=1}^{d_1+d_2} \sum_{j=1}^{N_i} \dot{\eta}_{ijr} \sum_{m'=1}^K x_{ijkm'}} & \text{if } 1 \leq r \leq S_1, \\ \frac{\sum_{i=d_1+1}^{d_1+d_2} \sum_{j=1}^{N_i} \dot{\eta}_{ijr} x_{ijkm}}{\sum_{i=d_1+1}^{d_1+d_2} \sum_{j=1}^{N_i} \dot{\eta}_{ijr} \sum_{m'=1}^K x_{ijkm'}} & \text{if } S_1 \leq r \leq S_1 + S_2. \end{cases} \quad (3.7)$$

CHAPTER 4 RESULTS

4.1 Hyper-Parameter Selection

Several different S_1/S_2 combinations were tested on the 13 book Pauline Corpus plus Hebrews (unknown authorship but attributed to Paul) and 1 John (known non-Pauline), both of which act as different types of controls. Initialization probabilities π for S_1 and S_2 as groups were determined by the proportion of sentences in the tested corpus that came from d_1 books and d_2 books respectively. In the thirteen book Pauline corpus, $\frac{\sum_{i=1}^7 N_i}{\sum_{i=1}^{13} N_i} = 0.683$ and $\frac{\sum_{i=8}^{15} N_i}{\sum_{i=1}^{13} N_i} = 0.317$, which when rounded for simplicity results in $\sum_{\gamma=1}^{S_1} \pi_\gamma = 0.7$ for S_1 styles and $\sum_{\gamma=R_1+1}^R \pi_\gamma = 0.3$ for S_2 styles. When either S_1 or S_2 contained two or more styles, their composite probability was allocated proportionally to create primary, secondary, and where necessary, tertiary styles.

All BICs for the tested models had very narrow ranges, so no individual model stood out as the obviously most superior. To provide clarity in the face of these ambiguities, we will use the exploratory model ($S_1 = 1, S_2 = 1$). As a binary classification of Pauline vs. Non-Pauline, ($S_1 = 1, S_2 = 1$) detects most simply the fundamental division that the final model will seek to detect greater complexity. The simplicity of ($S_1 = 1, S_2 = 1$), then, supplies a useful framework for interpreting these more complex, and potentially less clear, models. The final model ($S_1 = 4, S_2 = 2$) has four possible Pauline styles that can account for multiple clusters of books or sections of books. Two S_2 styles will permit the model to distinguish between multiple types of non-Pauline material, since at least known non-Pauline work (1 John) will be included in the model.

4.1.1 Nomenclature

When discussing styles, Pauline styles will be labeled PS and non-Pauline styles NPS. Each style designation has subscripts that differentiate amongst multiple styles in the same category. For example, in the ($S_1 = 1, S_2 = 1$) model, PS₁ refers Pauline style one, which is the lone Pauline style. NPS₁ refers to the lone non-Pauline style. In the ($S_1 = 4, S_2 = 2$) model, PS₁ through PS₄ refer to the four Pauline factor levels, and NPS₁ and NPS₂ refer to the two

non-Pauline styles.

4.2 Style Classifications

The linear progression of sentence classifications in constructed sequence plots (Figure 4.1 through Figure 4.6) provides a visual representation of the book's underlying style or styles, and thereby becomes a tool to make inferences regarding its authorship. In these sequence plots, the x -axis marks the sentences for the corresponding book on the y -axis. Each sentence is colored according to the final classification given it by the EM algorithm. For example, if the first sentence of Galatians were classified as Style 1, it might be colored sage green; but if it were classified as Style 2, it might be colored purple.

4.2.1 $S_1=1, S_2=1$ — 13 Book Claimed Pauline Corpus

An $(S_1 = 1, S_2 = 1)$ model has value for initial data exploration because its binary classification of Pauline vs. Non-Pauline is the fundamental division all subsequent models seek to detect, so its results supply a useful framework for interpreting more complex models. Before introducing Hebrews and 1 John as test books, we first examine the 13 book claimed Pauline corpus under the $(S_1 = 1, S_2 = 1)$ hyperparameter set to confirm the underlying classification patterns of disputed books in the absence of controls. Indeed, the $(S_1 = 1, S_2 = 1)$ model produced two clusters of books. The first cluster pattern (Figure 4.1) assigned nearly all non-Pauline sentences to 1 Timothy and Titus, hereafter called **Solution A***.

The second classification pattern that emerged (Figure 4.2) assigned nearly all non-Pauline sentences to Ephesians, Colossians, and 2 Thessalonians; or more particularly, the first halves of Ephesians Colossians, and 2 Thessalonians, hereafter called **Solution B***.

These two base classification patterns of books, Solution A* and Solution B*, match closely the clusters of books long noted by New Testament scholars as the Deutero-Pauline Epistles and the Pastoral Epistles, but with the curious omission of 2 Timothy from the Pastoral Epistles.

Twenty-three of the twenty-five solutions for $(S_1 = 1, S_2 = 1)$ on the 13 Book Pauline

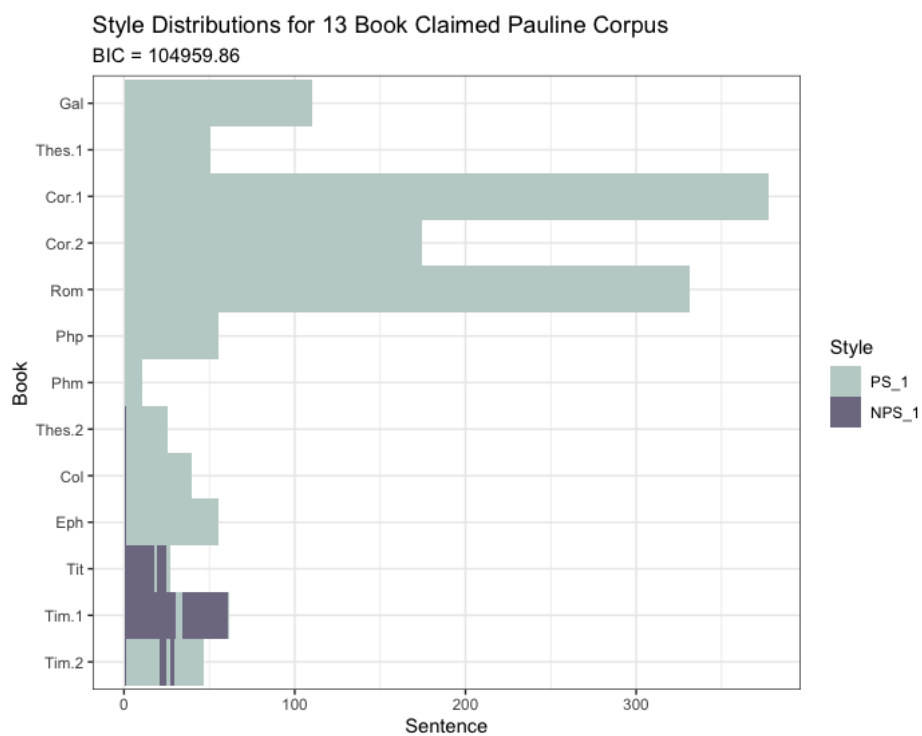


Figure 4.1 Solution A* (S1=1, S2=1)

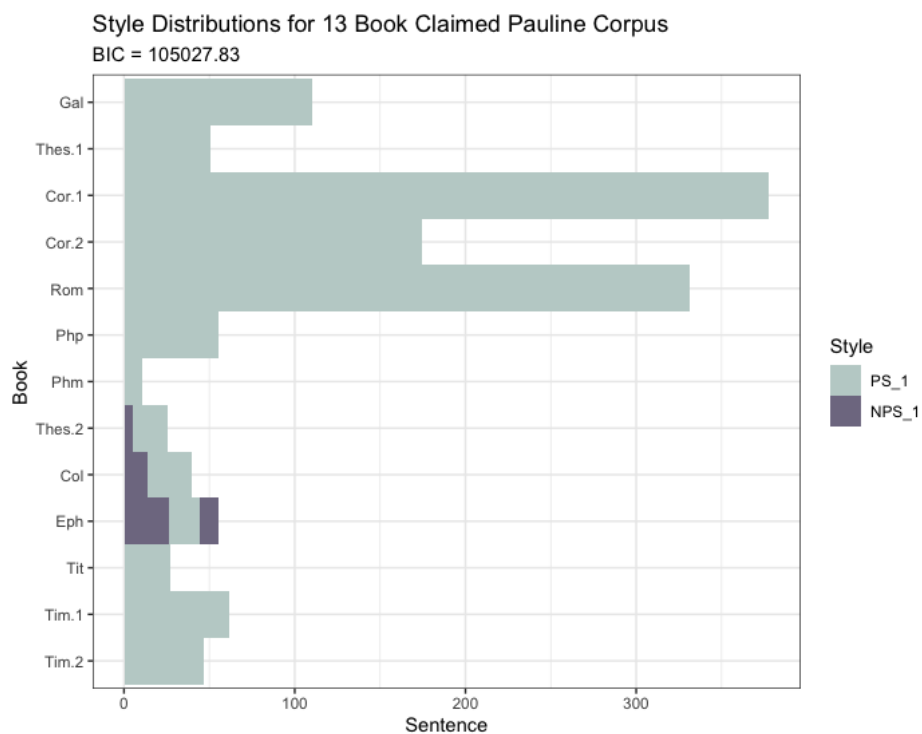


Figure 4.2 Solution B* (S1=1, S2=1)

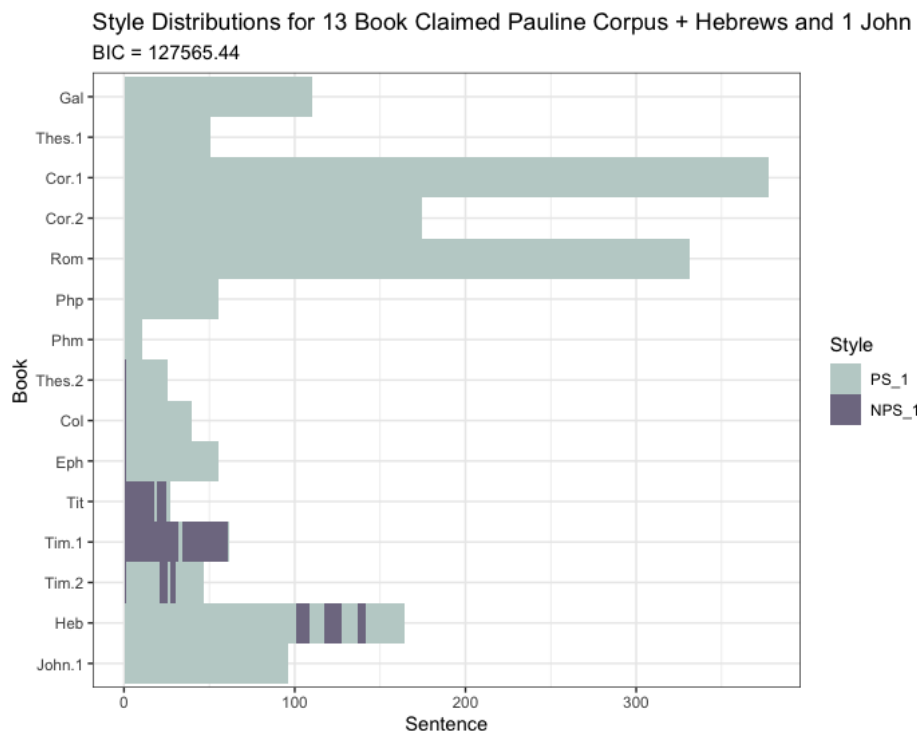


Figure 4.3 Solution A for (S1=1, S2=1) with Hebrews and 1 John

Corpus had BIC values to be considered viable. Of those twenty-three, six (26.1%) supported Solution A* and ranged in BIC from 104959.9 to 105003.9 (difference of 45). The remaining seventeen (73.9%) supported Solution B* and ranged in BIC from 105027.8 to 105034.5 (difference of 6.7). ID assignments within both groups were nearly identical.

Having confirmed the underlying clusters within the Pauline corpus, we now introduce Hebrews and 1 John as test books.

4.2.2 S1=1, S2=1 — 13 Book Claimed Pauline Corpus + Hebrews and 1 John

The complete test corpus, which includes all thirteen claimed Pauline books plus Hebrews and 1 John, produced patterns similar to those observed in Solution A* and Solution B*. On account of their similarities, these two new solutions will be referred to as **Solution A** (Figure 4.3) and **Solution B** (Figure 4.4). Both solutions classified small portions of Hebrews as non-Pauline, and in Solution B, 1 John was classified as non-Pauline; in Solution A, it was Pauline. 2 Timothy also continued to align more closely with the Undisputed Seven than with 1 Timothy and Titus.

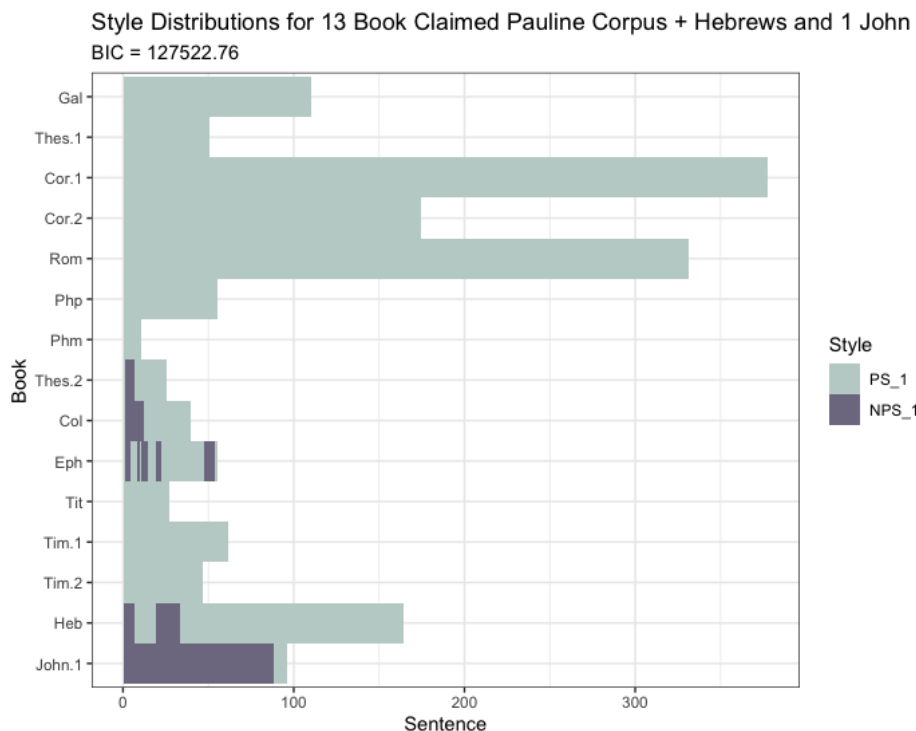


Figure 4.4 Solution B for ($S_1=1$, $S_2=1$) with Hebrews and 1 John

Twenty-two of the twenty-five solutions for ($S_1 = 1$, $S_2 = 1$) on the 13 Book Pauline Corpus + Hebrews and 1 John had BIC values to be considered viable. Three of those twenty-two (13.6%) supported Solution A and the remaining nineteen (86.4%) supported Solution B. Solution A's BIC values ranged from 127565.4 to 127567.1 (difference of 34.3), and Solution B's from 127522.8 to 127525.4 (difference of 2.6). Once again, id assignments within both groups were nearly identical. Therefore, it seems safe to conclude that ($S_1 = 1$, $S_2 = 1$) produces, on average, one of two different solutions: a dominant Solution B and a minority Solution A.

It makes sense, given the binary classification of ($S_1 = 1$, $S_2 = 1$), that Hebrews, as a book long attributed to Paul, should be classified as mostly Pauline; but 1 John, written by a confirmed non-Pauline author, is not always classified as non-Pauline (cf. Figure 4.3). Instead, it tends to fall within the same classification as Ephesians, Colossians, and 2 Thessalonians, revealing some classification dependency between them and 1 John. None of the best-performing solutions supported all six books—Ephesians, Colossians, 2 Thessalonians, 1 Timothy, Titus, and 1 John—as non-Pauline, a hypothetical “Solution X”. Nor did any of the best-performing models

Table 4.1 POS transition probabilities for Solution A (S1=1, S2=1)

	Style	AD	I	N	P	RA	V
AD	PS ₁	0.09	0.24	0.15	0.09	0.11	0.32
	NPS ₁	0.21	0.18	0.23	0.04	0.06	0.28
I	PS ₁	0.16	0.11	0.23	0.14	0.21	0.14
	NPS ₁	0.21	0.05	0.41	0.08	0.12	0.13
N	PS ₁	0.11	0.27	0.13	0.14	0.17	0.18
	NPS ₁	0.16	0.27	0.22	0.06	0.10	0.18
P	PS ₁	0.13	0.30	0.11	0.05	0.10	0.31
	NPS ₁	0.14	0.24	0.17	0.04	0.08	0.33
RA	PS ₁	0.13	0.08	0.66	0.02	0.01	0.10
	NPS ₁	0.21	0.11	0.52	0.01	0.03	0.12
V	PS ₁	0.10	0.39	0.10	0.15	0.18	0.07
	NPS ₁	0.15	0.35	0.13	0.09	0.15	0.14

classify only 1 John as non-Pauline and the remaining five disputed books as Pauline, a hypothetical “Solution Y”. We might describe the model’s apparent classification logic with three logical operators, each standing for some classification choice of non-Pauline material:

- A — Timothy and Titus.
- B — The first halves of Ephesians, Colossians, and 2 Thessalonians.
- J — 1 John.

Combined, they form the premise $A \vee (B \iff J)$, where $(B \iff J)$ is a biconditional that classifies either both or neither term as non-Pauline. Thus, if the model concludes $\neg A$, deciding that 1 Timothy and Titus are *not* non-Pauline (i.e. Pauline), then by implication both B and J are non-Pauline. Were the model to conclude either $\neg B$ or $\neg J$, then by implication A would be non-Pauline. That fact is made all the more striking by how markedly different 1 John’s (J) prose style is compared to Ephesians, Colossians, and 2 Thessalonians (B).

Table 4.1 contains the transition probabilities PS₁ and NPS₁ under Solution A. PS₁ is highlighted in blue, NPS₁ in orange.

Four significant differences between PS₁ and NPS₁ transition probabilities appear in Solution A. PS₁ has higher transition probabilities for AD - AD , I - N , and N - N than NPS₁, whereas

PS_1 has a higher transition probability for $RA-N$. Because $AD-AD$ and $I-N$ contain composite POS categories (cf. Table 4.1), it is difficult to know which particular transition or transitions are taking place.

Transitions $RA-N$ and $N-N$ are more straightforward as consecutive nouns or nouns immediately preceded by their definite article. For $N-N$, this could be either nouns in a list, or else one noun modifying another, perhaps in a genitive construction. Greek has twenty-four versions of the definite article, one for each of its gender-number-case combinations; and although a noun is not required to be immediately preceded by a definite article, the presence of the definite article serves multiple grammatical functions. Among other uses, it can be used to distinguish between subject and predicate in cases where it might be ambiguous, to emphasize a specific noun, to single out a specific object from its general class, or in reference to monadic objects. Sometimes, it is used or omitted for aural reasons and how the sentence sounds to the ear. Most often, an author will pair a noun with its definite article, as even PS_1 shows. While the difference between PS_1 and NPS_1 is significant in absolute terms, it is much smaller in relative terms compared to the difference in $N-N$ transitions. A higher $RA-N$ transition could indicate a more consistent or habitual use of the definite article, but not much more can be said apart from more specific textual data.

Solution B has three of the same differences Solution A has— $N-N$, $I-N$, and $RA-N$ —and adds $N-RA$ (Table 4.2). Since definite articles (RA) do not come after their nouns, these definite articles almost certainly belong to *next* noun, one that modifies or relates to the first. We can infer this from both Greek grammar and a typical construction in the genitive case, where one noun expresses a relationship to another, and also the high transition probability of $RA-N$ in NPS_2 (0.77). Most cases are likely a length-3 sequence $N-RA-N$.

But unlike Solution A, Solution B inverts which style has the greater probability, including for $N-RA$, converting Solution A's defining features of Pauline style (PS_1) into defining features of non-Pauline style (NPS_2). What in particular has driven these probability differences is unclear since the $(S_1 = 1, S_2 = 1)$ hyperparameter set is, by design, a sword rather than a scalpel.

Table 4.2 POS transition probabilities for Solution B ($S_1=1$, $S_2=1$)

	Style	AD	I	N	P	RA	V
AD	PS ₁	0.11	0.24	0.16	0.08	0.10	0.31
	NPS ₁	0.06	0.23	0.14	0.08	0.15	0.34
I	PS ₁	0.17	0.11	0.26	0.13	0.19	0.14
	NPS ₁	0.12	0.08	0.16	0.21	0.28	0.15
N	PS ₁	0.12	0.27	0.15	0.13	0.15	0.19
	NPS ₁	0.07	0.26	0.06	0.19	0.25	0.15
P	PS ₁	0.14	0.29	0.12	0.05	0.10	0.30
	NPS ₁	0.10	0.33	0.08	0.05	0.08	0.35
RA	PS ₁	0.15	0.09	0.62	0.02	0.02	0.10
	NPS ₁	0.11	0.04	0.77	0.00	0.00	0.09
V	PS ₁	0.11	0.38	0.11	0.14	0.17	0.08
	NPS ₁	0.08	0.40	0.06	0.15	0.26	0.06

Solution	Frobenius Distance	Absolute differences greater than 0.1
Solution A	0.15	4
Solution B	0.11	3

Table 4.3 Comparison between transition matrices: PS1 vs. NPS1

The Frobenius Distance between the styles in each solution is shown in Table 4.3, along with the number of differences between styles with an absolute value greater than 0.1. Solution A's Frobenius Distance is slightly larger, likely due to the large AD-AD transition difference, which does not appear in Solution B.

In sum, the initial examination of ($S_1 = 1$, $S_2 = 1$) presents three insights that will be seen more clearly, and become more important, in the final model:

- First, the two different solutions: Solution A and Solution B.
- Second, the front-loaded non-Pauline sentences in Solution B's Deutero-Pauline epistles.

Third, the large differences between styles for N - N and RA - N .

4.2.3 $S_1=4$, $S_2=2$

Among the higher-order models, ($S_1 = 4$, $S_2 = 2$) was chosen by the lowest-BIC criteria. In addition to the BIC, we added a minimum cluster-size condition of 50 for each style as a second criterion for valid solutions. The best-performing solution under ($S_1 = 4$, $S_2 = 2$) had

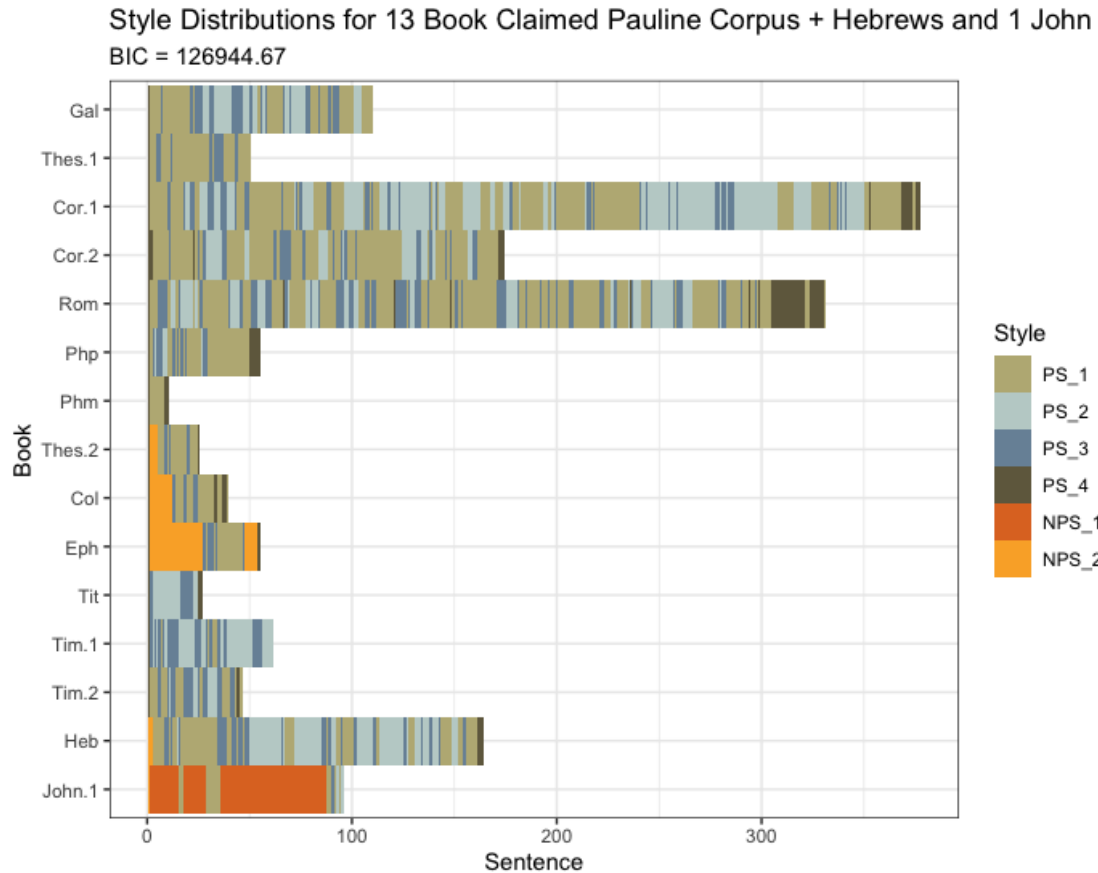


Figure 4.5 Solution C ($S_1=4$, $S_2=2$)

parallels that supported Solution B. We will call this solution **Solution C** (Figure 4.5). Based on Solution C's sentence classifications, we can describe the 15 book corpus as containing four groups of books:

1. The Undisputed Seven + 2 Timothy + Hebrews.
2. 1 Timothy and Titus.
3. Ephesians, Colossians, 2 Thessalonians (again, only their first halves).
4. 1 John.

($S_1 = 4$, $S_2 = 2$) likewise classified 2 Timothy as bearing more resemblance to the Undisputed Seven and Hebrews than to either of the two other Pastoral Epistles. All nine of these books share three characteristics: (1) a high proportion of the main Pauline style PS_1 , (2) smaller

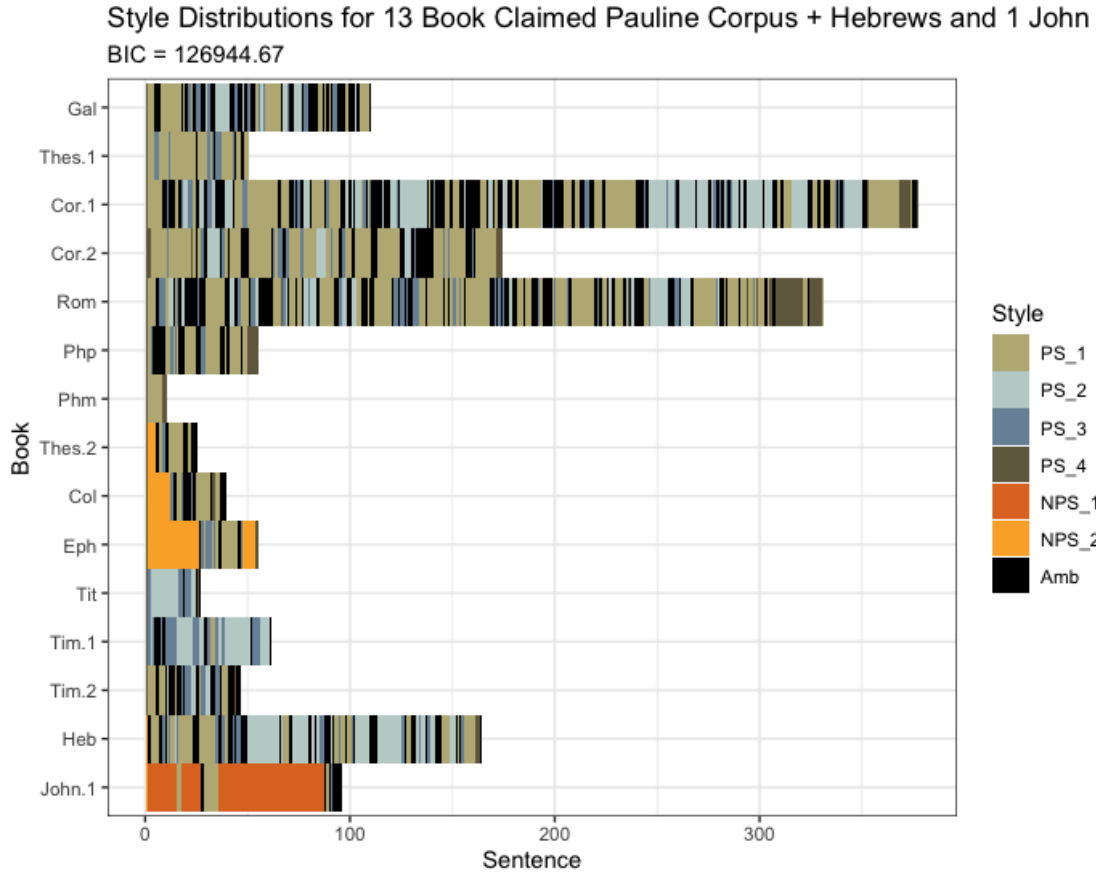


Figure 4.6 Solution C with ambiguities ($S_1=4$, $S_2=2$)

proportions of the secondary style, and (3) a smallest proportion of minority style PS_3 . Ten of thirteen claimed Pauline books close with at least several sentences of PS_4 , some with higher concentrations than others. Galatians, 1 Thessalonians, and 1 Timothy had no PS_4 sentence classifications in their closings. ($S_1 = 4$, $S_2 = 2$) also refined the results from ($S_1 = 1$, $S_2 = 1$) by distinguishing 1 John (NPS_1) and Deutero-Pauline epistles (NPS_2) as being *different* non-Pauline styles. Finally, it classified 1 Timothy and Titus as being composed in a *secondary* Pauline style (PS_2), related to but distinct from the Undisputed Seven.

But with six different styles being tested, the chances of ambiguous sentences (*viz.* sentences that have at least two styles with a 0.3 posterior probability or higher) increase, so we must also account for which books and which styles have the highest concentration of ambiguities (Figure 4.6). Ambiguous sentences are marked in black.

From Figure 4.6, it appears that the vast majority of ambiguous sentence classifications occur in the Undisputed Seven and 2 Timothy, and within those books they occur most often in the main Pauline style PS_1 and minority Pauline style PS_2 . NPS_1 , NPS_2 , and the PS_2 in 1 Timothy and Titus are left intact.

The POS transition probabilities (Table 4.4) for $(S_1 = 4, S_2 = 2)$ display the same differences as the exploratory model $(S_1 = 1, S_2 = 1)$. Interpreting these differences requires that we examine the underlying text where the styles in question are found. Each shall be expounded in turn, to the extent that inferences can be made.

1. $NPS_1 : AD - V$. AD is a composite category, which makes inference difficult. But adverbs are a rather broad category in Greek and quite fluid in their placement, yet that none of the other thirty-five transition probabilities from AD are greater than 0.34 is striking. A transition probability of 0.5, then, for whatever reason, would be consistent with the repetitive style of 1 John.
2. $PS_4 : I - N$. As was noted earlier, PS_4 appears almost exclusively in the epistles' final greetings, which provides strong evidence that PS_4 corresponds to the specific content of those sections, namely personal greetings. This is evidenced by three points. First, the heaviest concentration of PS_4 comes at the end of Romans, whose entire last chapter is a long list of personal greetings Paul sends to those in Rome. Second, those epistles that have no PS_4 —Galatians, 1 Thessalonians, and 1 Timothy—are a mixture of undisputed and disputed books, but they share one thing in common: they have no segment of final greetings. Third, Hebrews, though not part of the 13 book Pauline corpus, contains a final greeting classified as PS_4 . Yet because I is a composite category, knowing what this transition reflects grammatically or stylistically is somewhat unclear.
3. $PS_4 : N - N$. PS_4 has the highest transition probability of $N-N$, which possibly arises from the lists of names, being proper nouns, in the personal greetings. In Romans 16:14, for example, Paul names five consecutive men to greet: ἀσπάσασθε Ἀσύγκριτον Φλέγοντα

Table 4.4 POS transition probabilities for Solution B (S1=4, S2=2)

	Style	AD	I	N	P	RA	V
AD	PS ₁	0.10	0.24	0.13	0.12	0.10	0.32
	PS ₂	0.15	0.18	0.18	0.04	0.11	0.33
	PS ₃	0.08	0.34	0.17	0.06	0.06	0.29
	PS ₄	0.08	0.20	0.25	0.21	0.20	0.05
	NPS ₁	0.06	0.13	0.10	0.04	0.17	0.50
	NPS ₂	0.05	0.29	0.19	0.11	0.14	0.23
I	PS ₁	0.16	0.11	0.17	0.18	0.21	0.17
	PS ₂	0.20	0.14	0.19	0.09	0.26	0.12
	PS ₃	0.15	0.10	0.44	0.09	0.09	0.13
	PS ₄	0.08	0.01	0.65	0.05	0.20	0.01
	NPS ₁	0.10	0.10	0.07	0.27	0.32	0.15
	NPS ₂	0.14	0.06	0.23	0.19	0.25	0.14
N	PS ₁	0.08	0.25	0.19	0.20	0.18	0.19
	PS ₂	0.17	0.23	0.12	0.09	0.15	0.25
	PS ₃	0.13	0.24	0.22	0.05	0.08	0.18
	PS ₄	0.06	0.28	0.31	0.14	0.20	0.01
	NPS ₁	0.13	0.22	0.05	0.20	0.15	0.24
	NPS ₂	0.05	0.31	0.06	0.16	0.31	0.10
P	PS ₁	0.13	0.30	0.11	0.06	0.10	0.30
	PS ₂	0.19	0.23	0.15	0.04	0.07	0.32
	PS ₃	0.11	0.32	0.12	0.04	0.09	0.32
	PS ₄	0.09	0.46	0.22	0.03	0.10	0.11
	NPS ₁	0.12	0.29	0.03	0.04	0.08	0.44
	NPS ₂	0.10	0.33	0.09	0.07	0.09	0.32
RA	PS ₁	0.12	0.07	0.69	0.02	0.00	0.09
	PS ₂	0.20	0.07	0.57	0.02	0.03	0.11
	PS ₃	0.10	0.19	0.57	0.02	0.02	0.09
	PS ₄	0.13	0.13	0.57	0.00	0.00	0.16
	NPS ₁	0.07	0.03	0.79	0.00	0.00	0.11
	NPS ₂	0.13	0.04	0.76	0.00	0.01	0.07
V	PS ₁	0.09	0.42	0.07	0.19	0.15	0.08
	PS ₂	0.15	0.35	0.14	0.07	0.22	0.08
	PS ₃	0.12	0.36	0.17	0.12	0.15	0.10
	PS ₄	0.06	0.29	0.17	0.36	0.12	0.00
	NPS ₁	0.06	0.42	0.04	0.14	0.31	0.03
	NPS ₂	0.08	0.38	0.08	0.13	0.24	0.08

Ἐρμῆν Πατροβᾶν Ἐρμᾶν (“Greet Asyncritus, Phlegon, Hermes, Patrobas, Hermas”). A second example occurs at the end of Philemon in Philemon 1:24, where Paul names four men to whom he extends greetings: Μάρκος, Ἀρίσταρχος, Δημᾶς, Λουκᾶς (“[Greet] Mark, Aristarchus, Demas, Lucas”). Other textual causes for the high *N-N* transition likely exist as well, but are not as easily discerned.

4. $PS_4 : P - I$. PS_4 also has the highest proportion of *P-I* at 0.46. Again, the composite I category makes precise inferences difficult, but different iterations of the prepositional phrase “in Christ” are used to describe many of the individuals mentioned in these personal greetings. For example, in Romans 16:9, Paul writes, ἀσπᾶσασθε Οὐρβανὸν τὸν συνεργὸν ἡμῶν ἐν Χριστῷ. In a translation following Greek word order as much as possible, this would read, “Greet Urbanus, the fellow servant of ours in Christ”, with ἡμῶν (possessive pronoun, *P*) followed by ἐν (preposition, *I*). Elsewhere, in 1 Corinthians 16:19, Paul writes, ἀσπάζεται ὑμᾶς ἐν κυρίῳ πολλὰ Ἀκύλας καὶ Πρίσκα. (A close word order translation would be, “Greet you in the Lord heartily Aquila and Priscilla.”). As in the previous example, ὑμᾶς (object pronoun, *P*) precedes ἐν (preposition, *I*). These types of phrases, in this construction, appear throughout the personal greetings.
5. $NPS_1, NPS_2 : RA - N$. The more frequent use of the definite article could have any number of interpretations, some clearer than others, a fact further obscured by two things. First, the already high transition probabilities for the other four Pauline styles (0.57, 0.57, 0.69, and 0.57). NPS_1 material in 1 John is of a very different nature and subject matter than NPS_2 material in Ephesians, Colossians, and 2 Thessalonians, making it quite likely that the more frequent *RA-N* transition probability comes about for different reasons in each. Second, in terms of total words, NPS_1 and NPS_2 cover more material—and more varied material at that—than PS_4 , so the reasons given for the higher *RA-N* probability would, of necessity, be more general.
6. $NPS_2 : N - RA$. Reasons for this higher transition probability would be the same as that

given for the same phenomenon in $(S_1 = 1, S_2 = 1)$. As grammatical case is beyond the scope of this thesis, only more tentative, speculative conclusions can be reached from the higher *N-RA* transition probability, such as more frequent genitive constructions. Being concentrated in NPS_2 , however, would narrow the scope for any further quantitative or qualitative study.

The Frobenius distance from NPS_2 to the main Pauline style of PS_1 is the smallest distance of any combination, perhaps indicating a very subtle difference between the two. PS_2 , the secondary Pauline style contained in 1 Timothy and Titus, is the second-smallest distance and the second-closest to the main Pauline style PS_1 .

In terms of the magnitude of differences, Table 4.5 counts the total number of transition probability differences greater than 0.1. By this calculation as well, PS₄ is the most distinct style,

	PS₁	PS₂	PS₃	PS₄	NPS₁	NPS₂
PS₁	0					
PS₂	3	0				
PS₃	7	8	0			
PS₄	13	14	13	0		
NPS₁	9	7	15	18	0	
NPS₂	8	1	7	13	6	0

Table 4.5 Transition probability differences greater than 0.1

followed by NPS₁. Incidentally, each style has a POS category whose *initial* probability is near 0.5, far higher than all other initial probabilities in other styles. PS₄'s initial probability for *V* (0.507) comes from these sentences most often beginning with the command ἀσπᾶσθε (“Greet”). NPS₁'s initial probability for *I* (0.498) demonstrates 1 John's repetitive style, as nearly 30% of its sentences begin with either the conjunction καὶ (“and”) or the preposition ἐν (“in”) alone. That the model's classification choices correspond to lexical patterns readily observed in the text further supports not only the model's overall credibility but also the particular theory advanced earlier regarding the relationship between style and subject matter.

The distinct transition probabilities for PS₄ rule strongly in favor of Solution B's classification choices. Under ($S_1 = 1, S_2 = 1$), Solution A characterized NPS₁ as having a higher transition probabilities of *N-N*, *I-N*, and *P-I*, Solution B the inverse of that. ($S_1 = 4, S_2 = 2$), as a more precise model, parsed Solution B's characteristics into PS₄, the most distinct Pauline style among the six.

CHAPTER 5 INTERPRETATIONS

Before interpreting the results, it would be helpful to state the common observations yielded by the exploratory model ($S_1 = 1, S_2 = 1$) and the final model ($S_1 = 4, S_2 = 2$).

1. Despite being a Pastoral Epistle, 2 Timothy most often appeared indistinguishable from the Undisputed Seven.
2. The remaining disputed Pauline books formed two clusters, one containing Ephesians, Colossians, and 2 Thessalonians; the other 1 Timothy and Titus.
3. No solution classified all six disputed books, or even the six minus 2 Timothy, as non-Pauline.
4. No solution classified all six as Pauline either.
5. High transition probabilities for *N-N*, *RA-N*, and *N-RA* differentiated Pauline from non-Pauline styles, though which way depended on whether the results came from Solution A or Solution B.

Any one of these points would be inconclusive in and of itself; but together, they each contribute one piece to the greater whole, which we will now consider.

5.1 PS₄ and Subject Matter-Dependency

Over the centuries, theologians, publishers, and translators of the Bible have organized the text of Biblical books through a variety of typesetting features, adding chapters (13th century), verses (16th century), paragraphs (19th century), and then section headings (20th century). Features such as these, much like in any other book, outline the thematic and rhetorical structure of the books, and help the reader better follow what they are reading. Verses are the most granular text-break, followed by paragraphs, sections (or “pericopes”), and finally chapters. While not original to the text, there is widespread consensus about where these textual breaks should occur, with more major text divisions (e.g. chapters, sections/pericopes) being more uniform across

Book	Sentence for start of PS ₄ classifications	Chapter-Verse for start of PS ₄ classifications	Chapter-Verse for Personal Greetings
Romans	305	16:3	16:3
1 Corinthians	369	16:19	16:19
2 Corinthians	322	13:12	13:11
Ephesians	54	6:23	6:21
Philippians	50	4:19	4:21
Colossians	33	4:14	4:7
2 Thessalonians	25	3:17	3:16
Titus	25	3:15	3:15
2 Timothy	44	4:21	4:19
Philemon	9	1:23	1:23

Table 5.1 Verse-breaks for personal greetings

translations than more subtle divisions (e.g. paragraphs, verses), the latter depending largely on translation.

Classifications for PS₄ track closely with obvious text-breaks. The following table (Table 5.1) shows the sentence at which a book transitions to PS₄ and where the section of personal greetings begins, marked almost always by the Greek imperative ἀσπάσασθιε (“Greet”).

Galatians, 1 Thessalonians, and 1 Timothy have no section of personal greetings, and so are not listed in the table.

Four of the ten line up exactly with the transition to personal greetings. If the ambiguous sentences at the end of 2 Timothy are accounted for, 2 Timothy would line up exactly as well. The remainder, save Colossians, all fall within two verses, or one to two sentences, from the section. These passages also were the most likely to have been physically written by Paul’s own hand and without any collaboration or revision from his amanuensis (*i.e.* a scribal secretary). This acted as both a personal connection to the recipients and also a seal of the letter’s authenticity. We know from two undisputed epistles that Paul practiced this (cf. Galatians 6:11, Philippians 1:19), and two disputed epistles (cf. 2 Thessalonians 3:17 and Colossians 4:18) also bear witness to the same. That the classification as PS₄ almost universally implies personal greeting, and *vice versa*, validates that some sort of dependency exists between style and subject matter in the Pauline

Table 5.2 Summary data on sentence length by style classification

Style	Total	Word Count				
		Quartile 1	Median	Quartile 3	Mean	Standard Deviation
PS ₁	399	11	18	30	22.7	16.9
PS ₂	399	11	18	30	22.7	16.9
PS ₃	282	12	20	31.8	24.8	19.3
PS ₄	79	6.5	11	21.5	18	17.5
NPS ₁	77	15	21	27	22.5	10.3
NPS ₂	50	19	44.5	85.5	62.5	59.3

corpus.

Style and subject matter both might also be modified by rhetorical register. Furthermore, if rhetorical register, subject matter, or a combination of the two can create a shift from one Pauline style to another, it seems reasonable to think that they can create a shift from a non-Pauline to Pauline style. But if style has some dependency upon rhetorical register or subject matter, then a unique tone or unique subject could imagine a non-Pauline style where there is none. This appears to be the case with Ephesians, Colossians, and 2 Thessalonians because not all or most of their sentences are classified as non-Pauline, but only the first 30-50%. Far from random, such a pattern demonstrates an intentional choice about literary form and structure that likely produces different stylistic features.

Ephesians and Colossians have been described as being written in an “epideictic style”, which, amongst other things, was a high rhetorical form common in Asia Minor that often involves writing long, varied, and complex sentences.¹ Table 5.2 contains the data on sentence lengths for all six classification styles in ($S_1 = 4, S_2 = 2$). PS₄ has the smallest mean length, as greetings typically do not require as long of sentences compared to, say, doctrinal exposition. NPS₁ (1 John) has the smallest standard deviation in sentence length, a fact that accords with 1 John’s consistent and repetitive style. Finally, NPS₂, found only in the first halves of Ephesians, Colossians, and 2 Thessalonians, has by far the longest mean length and also the highest standard deviation in length. Those who doubt Paul’s authorship of these three books most often doubt it

¹Ben Witherington III, *The Letters to Philemon, the Colossians, and the Ephesians* (Eerdmans, 2007), 4-19.

Table 5.3 Posterior probabilities for Ephesians

Sentence	Style					
	PS ₁	PS ₂	PS ₃	PS ₄	NPS ₁	NPS ₂
21	0	0	0	0	0	1
22	0.007	0	0	0	0	0.993
23	0.107	0.001	0.023	0	0	0.869
24	0.117	0.038	0.015	0	0	0.83
25	0.092	0.075	0.028	0	0	0.804
26	0.155	0.071	0.013	0	0	0.761
27	0.17	0.072	0.056	0	0	0.702
28	0.022	0.001	0.831	0.004	0	0.144
29	0.713	0.048	0.166	0.001	0	0.072
30	0.189	0.001	0.774	0.002	0	0.034
31	0.004	0.014	0.982	0	0	0

on two grounds: sentence length and subject matter. The long sentences in those books, much longer than is typical in Paul's undisputed works, is often held to be inconsistent with Pauline style; and their content usually has but few direct thematic links to the undisputed letters. Yet ironically, both of these objections combine to *weaken* the argument against their Pauline authorship. For if style is in some degree dependent on subject matter, then those books' unique material has a distinct possibility of begetting a unique style. This becomes all the most probable in light of the epideictic rhetorical style used in Ephesians, Colossians, and possibly 2 Thessalonians, and also that the rest of the book is classified as Pauline.

This theory can be tested by examining the posterior probabilities for their sentence classifications, just prior to their transition from NPS₂ to the main Pauline style PS₂ (Table 5.3, Table 5.4, and Table 5.5). The posterior probabilities for NPS₂ in all three books follow a similar pattern. They begin at 1 or near 1 probability for NPS₂, and then decline before abruptly and permanently switching to the a Pauline style. The lone exception to the "permanent" switch is the tail end of Ephesians, which we will revisit. That decline and abrupt change speaks to a transition in the text, and when we see where these transitions take place in the text, they have similarly conspicuous break-points as did the PS₄ sentences. In this case, those divisions of the text are when the book shifts from doctrinal exposition to personal exhortation, from the indicative to the

Table 5.4 Posterior probabilities for Colossians

Sentence	Style					
	PS ₁	PS ₂	PS ₃	PS ₄	NPS ₁	NPS ₂
9	0	0	0	0	0	1
10	0	0	0.039	0	0	0.96
11	0.029	0	0	0	0	0.971
12	0.069	0	0.003	0	0	0.928
13	0.001	0	0.964	0	0	0.034
14	0.513	0.466	0	0	0.014	0.007
15	0.554	0.434	0.007	0	0	0.005

Table 5.5 Posterior probabilities for 2 Thessalonians

Sentence	Style					
	PS ₁	PS ₂	PS ₃	PS ₄	NPS ₁	NPS ₂
2	0.013	0	0	0	0	0.987
3	0.019	0	0	0	0	0.981
4	0.151	0	0	0	0	0.849
5	0.196	0	0	0	0	0.804
6	0.671	0.001	0.058	0	0.016	0.254
7	0.848	0.063	0.028	0	0	0.061

imperative, a major theme in Pauline theology.² Most often, the doctrinal exposition forms the basis for the specific exhortations given, and hence why it is written first.

Table 5.6 lists the chapters and verses that mark the classification pattern noted in Ephesians, Colossians, and 2 Thessalonians, read from the English Standard Version (ESV), a modern Bible translation based on the Greek text used in this thesis. The probabilities mentioned in the table reflect the mean posterior probability of the dominant class. All but one classification point begins and ends on or almost on some major text division, be it the beginning of a chapter or section/pericope heading. (The lone exception, “Transition” for 2 Thessalonians, breaks at a new paragraph in some translations, and in the middle of that paragraph in others.) So improbable is this that the underlying subject matter of the text, combined with the rhetorical register, must dictate the style of the books’ doctrinal exposition. Returning to the lone exception of permanent

²Herman Ridderbos, *Paul: An Outline of His Theology* (Grand Rapids, 1997), 256.

Michael Parsons, “Being Precedes Act: Indicative and Imperative in Paul’s Writing”, *Evangelical Quarterly* 88, no. 2 (1988), 99-127.

	Exposition	Transition	Exhortation
Posterior probability of dominant class	Non-Pauline (1 or near 1)	Non-Pauline (0.7 to 0.8)	Pauline (probability near 0.9)
Ephesians	1:2-4:16, 6:9,10-6:20,21	4:17-4:31,32	5:1-6:9
Colossians	1:3-2:15	2:16-2:23	3:1-4:6
2 Thessalonians	1:3-1:12	2:1-2:4	2:5-3:15

Table 5.6 Verse breaks for style transitions, based on posterior probabilities

transition to Pauline material, Ephesians contains one section of NPS₂ near its end, in an elaborate conceit known as “The Armor of God” (Ephesians 6:10-20), singular in literary application to the book of Ephesians. Just like the other sections described previously, the classification of NPS₂ sentences extends almost exactly over the verses covered by The Armor of God, transiting to the PS₄ personal greetings before its close. In the end, thematic and literary variation proffers the simplest and most consistent explanation for differences among at least some Pauline styles, as well as between Pauline and ostensible (but superficial) non-Pauline styles.

Thus, arguments against these books’ authenticity err in their assumption that Pauline styles remain largely invariant to changes in rhetorical register or subject matter. If style did not depend in some large measure upon rhetorical register or subject matter, then we should expect a more random distribution of non-Pauline sentences throughout disputed books. We see the opposite, in fact. Such abnormal distributions point toward authorial intent and should be investigated accordingly. Their Pauline authorship should not be dismissed out of hand so hastily, notwithstanding even the arguments presented, for they contain still either a majority or large plurality of Pauline styles. At the very least, we can conclude that their non-Pauline character has been exaggerated.

5.2 Application to the Pastoral Epistles

Though 2 Timothy has always been considered one of the Pastoral Epistles, it clearly stands apart from 1 Timothy and Titus, and displays little if any affinity with their style classification, instead appearing indistinguishable from authentic Pauline letters. This observation comports with the work of various New Testament scholars, who have argued that the authorship

of 1 Timothy, 2 Timothy, and Titus must be judged individually rather than collectively, and who think that Paul wrote 2 Timothy.³ Unlike 1 Timothy and Titus, 2 Timothy contains no detailed instructions about ecclesiastical order, follows a different literary form, and is far more personal and intimate in its tone. Given these textual dissimilarities and the consistent linking of 1 Timothy and Titus over and against 2 Timothy, it appears that the Pastoral Epistles can be grouped by one of two criteria: similar recipients or similar content. The evidence presented above would incline toward the latter, and therefore restrict the category of Pastoral Epistles (at least for debates over authorship) to 1 Timothy and Titus only. As far as the null hypothesis of Pauline authorship is concerned, neither $(S_1 = 1, S_2 = 1)$ nor $(S_1 = 4, S_2 = 2)$ have provided sufficient evidence to reject it regarding 2 Timothy, so 2 Timothy's assumed Pauline authorship ought to be maintained.

Unlike PS_4 , and NPS_2 , there are few obvious cluster patterns in the Undisputed Seven for the secondary Pauline style PS_2 in 1 Timothy and Titus. But because rhetorical register and subject matter has been shown to influence style in the case of PS_4 and NPS_2 , we at least have *prima facie* evidence to think that they likely play some role in 1 Timothy and Titus, though what specifically might be hard to determine. Still, a few theories are in order.

One curious difference in the transition probabilities between the main Pauline style PS_1 and the secondary PS_2 in the Pastoral Epistles is the decreased likelihood of N-P transitions, noun-to-pronoun, in PS_2 . An examination of several passages gives credence to the idea that the decreased likelihood of N-P transitions could come from a lack of genitive or possessive pronouns relative to nouns. Indeed, Table 5.8 shows that the two most common styles in 1 Timothy and Titus, PS_2 and PS_3 , which make up between 80% and 90% of their sentences, have a much higher ratio of nouns to pronouns than PS_1 . Interestingly, although PS_4 has a higher noun to pronoun ratio than PS_1 , its N-P transition probability is higher, likely a consequence of the personal greetings. In the example from Romans 16:9 cited earlier (ἀσπάσασθε Οὐρβανὸν τὸν συνεργὸν ἡμῶν ἐν Χριστῷ), the noun συνεργὸν precedes the possessive pronoun ἡμῶν. Throughout the

³Jerome Murphy-O'Connor, *Paul: A Critical Life*, (Clarendon Press, 1996), 356–359. Jerome Murphy-O'Connor, *Paul the Letter-Writer: His World, His Options, His Skills* (The Liturgical Press, 1995), 48. Michael Prior, *Paul the Letter-Writer and the Second Letter to Timothy* (Bloomsbury Publishing, 1989).

	AD	I	N	P	RA	V
PS-1	0.121	0.232	0.189	0.137	0.137	0.184
PS-2	0.176	0.203	0.218	0.061	0.151	0.189
PS-3	0.124	0.264	0.289	0.072	0.088	0.163
PS-4	0.077	0.2	0.389	0.105	0.157	0.072
NPS-5	0.095	0.209	0.187	0.132	0.18	0.197
NPS-6	0.097	0.216	0.242	0.12	0.185	0.14

Table 5.7 Proportion for all POS in each style

personal greetings, Paul describes individuals as “our brother”, “my fellow servant”, etc.

The comparative scarcity of pronouns in PS₂ perhaps reflects a less intimate style suited to things and ideas, not persons. This could make sense of why Philemon, a letter written to a close friend of Paul’s asking him to forgive and free his slave Onesimus, has the highest proportion of pronouns of any book (0.183); and by contrast, why 1 Timothy, a letter concerned with matters of church government and administrative affairs, has the lowest proportion of pronouns (0.068). And although Philemon is only ten sentences long, leaving little room for substantive inference, it is worth noting that Philemon has no PS₂ or PS₃ sentences. If we accept that POS transition probabilities could differ by literary form as well, then Johnson’s theory that 1 Timothy and Titus represent a type of *mandata principis* could explain the divergence of those two books as well.

Although the explanations are less clear for 1 Timothy and Titus, unlike Ephesians, Colossians, and 2 Thessalonians, the Pastoral Epistles most often start from a heavily Pauline classification; and as was mentioned at the outset of this thesis, explaining continuity is more difficult than explaining discontinuity. Nevertheless, the theories presented give linguistically and thematically plausible explanations for what variations and phenomena do exist, explanations which have previously been advanced by New Testament scholars through more qualitative methods.⁴ The differences in POS transition probabilities fulfilled the two-fold criteria for validity stated in 4.1.3, being both significant and meaningful. All of these theories, moreover, address only internal evidence and explanatory variables within the text itself. They did not consider

⁴Johnson, *Letters to Paul’s Delegates* (Bloomsbury Academic Press, 1996), 11-12. Johnson notes that what many consider “non-Pauline vocabulary” is concentrated in passages concerned with specific heresies addressed only in the Pastoral Epistles.

external evidence and explanatory variables from outside the text. The universal reception of these epistles by the early church would add weight to their authenticity. In addition to this, we mentioned Paul's use of amanuenses in writing his epistles (cf. Romans 16:22). Amanuenses played important roles in letter-writing, acting as scribes, editors, co-authors, and even composers, depending on the circumstances. What role they might have played in shaping the language is difficult to say.⁵ Finally, almost all of Paul's letters also open by naming apostolic colleagues with whom he could possibly have coauthored the letter. Because such theories from external evidence cannot be conclusively tested against, they were not factored into the models, but their reasonable historical plausibility should further temper arguments about anomalies in the PE.

⁵E. Randolph Richards, *The Secretary in the Letters of Paul* (Mohr, 1991), 43-56; E. Randolph Richards, *Paul and First-Century Letter Writing Secretaries, Composition and Collection* (InterVarsity Press, 2004), 59-92.

CHAPTER 6 CONCLUSIONS

By taking New Testament scholarship into account, the results presented in this thesis demonstrate that disputed Pauline books, the Deutero-Pauline Epistles and the Pastoral Epistles, are best understood as Pauline works. For the Deutero-Pauline Epistles, their non-Pauline sentences—especially for Ephesians and Colossians—consistently map onto the indicative sections of the indicative-imperative structure in those letters, which have a unique combination of rhetorical tone and subject matter. Those two authorial choices likely result in sentences longer on average than those found in the Undisputed Seven, but rhetorical tone and unique subject matter cannot act as proxy variables for a non-Pauline because they cannot be tested against. Considering that only Ephesians contained a majority of NPS₂ (still only 58%) but had the strongest dependency on subject matter, it is reasonable to retain the null hypothesis for them.

For the Pastoral Epistles, the case is more straightforward, since they contained no non-Pauline styles in the final model. 2 Timothy appeared no different than the Undisputed Seven, even when multiple Pauline and non-Pauline parameters were added. One Pauline style (PS₄) and one non-Pauline style (NPS₂) had clear evidence of being dependent on a combination of rhetorical register and subject matter, making it stand to reason that other styles do as well. 1 Timothy and Titus share common purpose, tone, and subject matter, and are written in secondary Pauline styles that have POS transition probabilities with plausible links to those commonalities. We ought, therefore to conclude that the null hypothesis of Pauline authorship has insufficient evidence against it, and so retain the null hypothesis.

Debates over Pauline authorship of the Pastoral Epistles too often suffer from weak assumptions and superficial answers. Should attempts to provide statistical data on the subject fail to derive their methodologies from a thorough review of robust New Testament scholarship, they shall only exacerbate these faults, for their reach exceeds their grasp. And given enough uncritical reception, such flawed and deficient methods turn mantras into dogmas not easily dislodged from their place of privileged error.

In all likelihood, debates over the Pastoral Epistles will continue. But absent convincing

evidence for non-Pauline origin, those debates would be most productive if focused on integrating the Pastoral Epistles into the Pauline corpus, rather than still casting doubts on their Pauline authorship; doubts, it should be added, that are presupposed, but not proven. Sometimes, things just are what they appear to be, and in the case of the Pastoral Epistles, that means letters written by Paul to Timothy and Titus. That may not be novel, but then again, truth rarely is.

CHAPTER 7 BIBLIOGRAPHY

Baugh, S. M. *Ephesians: Evangelical Exegetical Commentary*. Lexham Press, 2016. Baum, L.E. T. Petrie, G. Soules, and N. Weiss. "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains." *The Annals of Mathematical Statistics* 41, no. 1 (1970): 164–171.

Beale, G.K. *Colossians and Philemon: A Paragraph-by-Paragraph Exegetical Evangelical Bible Commentary*. Baker Academic Press, 2019.

Dempster, A.P., N.M. Laird, and D.B. Rubin. "Maximum likelihood for incomplete data via the EM algorithm (with discussion)." *Journal of the Royal Statistical Society, Series B* 39, no. 1 (1977): 1–38.

Fee, Gordon. *The First and Second Letters to the Thessalonians (New International Commentary on the New Testament)*. Eerdmans, 2009.

Foster, Paul. "Who Wrote 2 Thessalonians? A Fresh Look at an Old Problem." *Journal for the Study of the New Testament* 35, no. 2 (2012): 150-175.
<https://doi.org/10.1177/0142064X12462654>.

Hoehner, Harold. *Ephesians: An Exegetical Commentary* (Baker Publishing Group, 2001). Johnson, Luke Timothy. *The First and Second Letters to Timothy: A New Translation with Introduction and Commentary*. Yale University Press, 2008.

Johnson, Luke Timothy. *Letters to Paul's Delegates*. Bloomsbury Academic Press, 1996.

Knight III, George W. *The Pastoral Epistles: A Commentary On the Greek Text*. Eerdmans, 1992.

Krueger, Michael J. "The Authenticity of 2 Peter", *Journal of Evangelical Theological Society* 42, no. 4 (1999): 645-671.

Lewis, C.S. *Christian Reflections*. Eerdmans, 2019.

Metzger, Bruce. "A Reconsideration of Certain Arguments Against the Pauline Authorship of the Pastoral Epistles", *The Expository Times*, 70, no. 3 (December 1958): 91-94.

Murphy-O'Connor, Jerome. *Paul: A Critical Life*. Clarendon Press, 1996.

Murphy-O'Connor, *Paul the Letter-Writer: His World, His Options, His Skills*. The Liturgical Press, 1995.

Neumann, Kenneth J. *The Authenticity of the Pauline Epistles in the Light of Stylostatistical Analysis*. Scholars Press, 1990.

Parsons, Michael. "Being Precedes Act: Indicative and Imperative in Paul's Writing", *Evangelical Quarterly* 88, no. 2 (1988): 99-127.

Prior, Michael. *Paul the Letter-Writer and the Second Letter to Timothy*. Bloomsbury Publishing, 1989.

Rabiner, L.R. "A tutorial on hidden Markov models and selected applications in speech recognition." *Proceedings of the IEEE* 77, no. 2 (1989): 257-286.

Richards, Randolph E. *Paul and First-Century Letter Writing Secretaries, Composition and Collection*. InterVarsity Press, 2004.

Richards, Randolph E. *The Secretary in the Letteres of Paul*. Mohr, 1991.

Ridderbos, Herman. *Paul: An Outline of His Theology*. Eerdmans, 1997).

Robinson, J.A.T. *Redating the New Testament*. Wipf and Stock, 2000.

Tooth, Sydney E. "Suddenness and Signs: The Eschatologies of 1 and 2 Thessalonians." PhD diss, University of Edinburgh, 2019.

Towner, Philip H. *The Letters to Timothy and Titus*. Eerdmans, 2006.

Witherington III, Ben. *The Letters to Philemon, the Colossians, and the Ephesians*. Eerdmans, 2007.