

**DISENTANGLING AUTHORSHIP AND GENRE IN THE GREEK NEW
TESTAMENT:
HISTORY, METHOD AND PRAXIS**

by

James A. Libby, B.S, M.Div.

**A dissertation submitted to
the faculty of McMaster Divinity College
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy (Christian Theology)**

**McMaster Divinity College
Hamilton, Ontario
2015**

Doctor of Philosophy (Christian Theology)
MCMASTER DIVINITY COLLEGE
Hamilton, Ontario

TITLE: Disentangling Authorship and Genre in the Greek New Testament: History, Method and Praxis

AUTHOR: James A. Libby

SUPERVISORS(S) Dr. Stanley E. Porter
Dr. Christopher D. Land

NUMBER OF PAGES: ix + 404



McMASTER DIVINITY COLLEGE

Upon the recommendation of an oral examining committee,

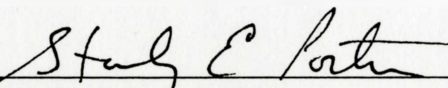
this dissertation by

James A. Libby

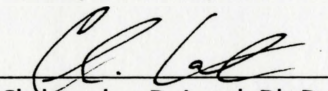
is hereby accepted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY (CHRISTIAN THEOLOGY)

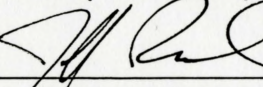
Primary Supervisor:


Stanley E. Porter, Ph.D.

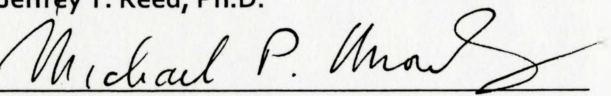
Secondary Supervisor:


Christopher D. Land, Ph.D.

External Examiner:


Jeffrey T. Reed, Ph.D.

Academic Dean Designate:


Michael P. Knowles, Th.D.

Date: April 8, 2015

ABSTRACT

“Disentangling Authorship and Genre in the Greek New Testament: History, Method and Praxis”

James A. Libby
McMaster Divinity College
Hamilton, Ontario
Doctor of Philosophy (Christian Theology), 2015

This is a work that explores linguistic style within the Greek New Testament and the extent to which its presumed causes (e.g. differences in authorship, genre, topic, subject matter, and the like) can be “disentangled.” Scarcely anyone who has perused the long history of authorship debates in the New Testament can mistake the profound implications such a “disentangling” would bring. Two motivations exist to revisit this issue: compelling recent findings in computational stylistics, and the sheer theological implications such a study may bring. Concerning the former motivation, earlier generations of scholars assigned, *de facto*, virtually any significant stylistic variation to authorship alone. The last thirty years of research outside NT studies, however, has demonstrated that more frequently than not, more of the total summed stylistic variation in mixed genre corpora is due to genre rather than authorship. What, indeed, would be the implications if those findings proved true of the GNT as well? First, and somewhat deconstructively, if the major proportion of stylistic variation in the GNT were found to be due to genre rather than authorship—or even close to it—any prior studies that had (1) failed to test for genre as a competing theory or (2) failed to remove genre as a covariate have almost certainly confounded genre with authorship. Second, and more constructively, if a convincing separation between what is commonly termed authorial variation and the various sources of sociolectic variation (which include component as genre/register, subject matter,

audience, and the like) can be achieved, such a thing would have broad and sweeping impact upon many *topoi* within New Testament scholarship. Not only would it influence the obvious suspects (i.e. the authorship of the Pastoral Epistles, the extent of the Pauline Canon, pseudepigraphy, the Synoptic Problem, and the like) it would also necessarily influence the current vigorous discourse in New Testament hermeneutics itself. My approach is threefold. First, the history of computational stylistics both within and outside NT studies will be reviewed. Second, as to method, an abductive approach, one that harnesses both Systemic Functional Linguistics and a variety of univariate and multivariate methods will be adopted. Third, that method will be exercised on the text of the Greek New Testament itself.

DEDICATION

There are in every life turning points. No sentinel announces them, they just come. For me, at least, these turning points have come from the examples of others. A kind word here, a question there, or most profoundly, the example of a life well-lived. This dissertation, so long delayed, is the product of those kinds of examples and those kinds of people.

I first wish to recall the example of Donald Burdick, my first professor of Greek at Denver Seminary who at the end of a long teaching career nonetheless came to sit down next to me (and next to a computer of all things!) to discover how that odd thing could help him learn more about the Greek New Testament that he so dearly loved.

Next, there is the example of Stanley E. Porter whose peaceful and unhurried demeanor, so different from my own, belies a breadth and depth of a scholarship that have caused many of his students—and most of all me—to marvel at how he does it.

Most importantly, there is my dear, dear, bride Tori, my patient life-companion of thirty three years. She has taught me more about love and servanthood than all of my teachers.

And last of all I dedicate this work to Dad, who saw the beginning of this project but not its end. You taught me wordlessly, by your example, to love learning.

Table of Contents

1	Introduction: Disentangling Authorship and Genre in the GNT	1
1.1	Introduction: The Tangled Dance of Authorship and Genre.....	1
1.2	Recent Advances in Computational Stylistics: Its Role in “Disentangling”.....	2
1.3	Implications: What if Similar Findings <i>Were</i> Verified in the GNT?.....	6
1.4	The Theological Potential of “Disentangling”: A Pastoral Example.....	7
1.5	Not So Fast: The Need for Theory	15
1.6	The Central Objective and Structure of This Work	19
2	Ground Clearing	21
2.1	Developing a Theory of Disentangling: Proposing an Abductive Approach.....	21
2.2	Articulating Our Starting Stances (Priors)	27
2.3	Gaps in the Current State of NT Computational Stylistics	31
2.4	Computational Stylistics: A Primer.....	36
2.5	Linguistics: A Primer	57
2.6	Before Beginning: Haven’t We Forgotten Something?	71
2.7	Remind Me Again: Why Is This Theology?	75
3	History	77
3.1	What Is an Experimental Design Review?.....	77
3.2	Period One: Quantitative Stylistics from Antiquity to Campbell	79
3.3	Period Two: Stylometric Descriptivism (1867–1935)	81
3.4	Milestone One: Modern Linguistics Emerges (1890s–1920)	93
3.5	Milestone Two: The Birth of Statistics and Quantitative Linguistics (1922–1935)	96
3.6	Period Three: Harps in Babylon: Description without Inference (1922–1960s)	100
3.7	Milestone Three: First Light: Mosteller, Wallace and Herdan (1950s–1966)	105
3.8	Period Four: Stylometry Amidst the Positivist Bacchanal (1930s–1980)	112
3.9	Milestone Four: The New Stylistics 1.0: Disentangling the Layers (Late 1980s)	131
3.10	Period Five: The Interregnum That Led to the Wild West (Mid-1980s–1990s).....	138
3.11	Period Six: The Mess (The 1990s–Mid-2000s)	155
3.12	Period Seven: The New Stylistics 2.0: Machine Learning (Mid-1990s–Current)	162
3.13	Milestone Five: Inviting Linguistics to the Dance (ca. 2005–Current).....	166
4	Developing the Design for Disentangling Authorship from Genre.....	168
4.1	Define the Research Scope.....	169
4.2	Implications from the Design Review.....	175

4.3	Assemble Implications from Current Large-Scale Work.....	179
4.4	Develop the Initial Three Dimensions of the Synthetic Experimental Design	182
4.5	Explore Technologies to Automate the Execution of the Experimental Design.....	194
4.6	Given This Experimental Design—What Should We Chose to Do First?	197
5	Defining Testable Language Measures, Authorship Theories and Genre Theories.....	199
5.1	Framing the First Cycle Through the Abductive Design (Chapters 6–9)	199
5.2	Developing Testable Language Measures.....	202
5.3	Developing Testable Authorship Theories.....	206
5.4	Developing Testable Genre Theories	210
6	Results Part I: Testing GNT Authorship and Genre Using Univariate Analysis	232
6.1	Preparing the GNT for Analysis.....	232
6.2	Interpreting Contingency Tables and Their Statistics: A Primer	235
6.3	Testing Bivariate Association, Direction and Significance	239
6.4	Bivariate Information Theory: Another Lens for Viewing Authorship and Genre.....	250
7	Results Part II: Disentangling GNT Authorship and Genre Using Multivariate Analysis..	257
7.1	Disentangling Authorship from Genre Using HLLA: A Primer.....	257
7.2	Hypothesis One: Is Authorship or Genre “Stronger” Overall?	262
7.3	Hypothesis Two: Is Authorship or Genre More Associated with <i>Infrequent Data</i> ?	264
7.4	Hypothesis Three: How Do the Strongest Theories Compare?	268
7.5	An Abductive Finding: An Authorship Pattern in the HLLA Deletion Statistics?.....	274
7.6	Disentangling Authorship from Genre Using HLLA: Conclusions.....	280
8	Results Part III: Visualizing NT Text Clustering Using Correspondence Analysis.....	286
8.1	Visualizing How the NT Texts Cluster: An EMVA Primer	286
8.2	The Experimental Design Used to Perform the Correspondence Analyses.....	289
8.3	An Exercise in Interpreting Correspondence Analysis	291
8.4	Visualizing Lexematic, Paradigmatic and Semantic Data Using CA	296
8.5	Using Euclidean Distance to Test How Well The Theories Fit the CA Data	301
9	Results Part IV: Exploring GNT Information Loss Via Information Theory	305
9.1	The History of Information Theory and Linguistics	306
9.2	Studying the GNT Using Information Theory: A Primer	310
9.3	Exploration One: Does Authorship or Genre Contain the Most Information?	316
9.4	Exploration Two: Which Language Fraction Encodes the Most Information?	319
9.5	Exploration Three: How Does TI Loss Differ by Language Measure in the GNT?	321
9.6	Conclusions: Initial Information Theory Explorations into Authorship and Genre.....	324

10	Summary Conclusions and Hermeneutical Reflections.....	325
10.1	Findings Summary	326
10.2	Implications for NT Studies	333
10.3	Implications in Terms of Methodology.....	338
10.4	Directions for Further Research	342
10.5	A Final Hermeneutical Reflection.....	344
	Appendix.....	348
	Bibliography	350

1 Introduction: Disentangling Authorship and Genre in the GNT

1.1 Introduction: The Tangled Dance of Authorship and Genre

This is a work that explores literary and linguistic style within the Greek New Testament (henceforth GNT) and the extent to which its presumed causes (e.g. differences in authorship, genre, subject matter, audience, and the like) can be “disentangled.” Scarcely anyone who has perused the long history of authorship debates in the New Testament,¹ especially regarding its still contested relationship to the praxis of Greco-Roman pseudepigraphy,² can mistake the profound implications such a “disentangling” would bring. The heart of the matter is this: can the obvious stylistic differences between texts³ in the GNT be convincingly demonstrated to arise from one cause (e.g. authorship, genre, etc.) rather than another? Or can they be shown, at least, to have arisen *more* from one cause than another? Such findings would not only inform the traditional domains of occasion and introduction in New Testament studies, they would markedly advance our understanding of early Christian origins. The problem? Historically, much scholarly capital has been spent attempting to quantitatively attach stylistic phenomena to this or

¹ While Erasmus doubted the authenticity of Ephesians as early as the sixteenth century, the pruning of the Pauline canon began in earnest in the nineteenth century. The sequence of this reduction and the scholars first responsible for it were as follows: 1 Timothy: J.E.C. Schmidt (1804) and Schleiermacher (1807); all three Pastorals: J.G. Eichhorn (1812); the entire Pauline Canon save the *Hauptbriefe*: F.C. Baur (1845); 2 Thessalonians and Colossians: Hilgenfeld (1875) and H.J. Holtzmann (1885). (Erasmus, *Annotaciones*; Schmidt, *Historisch-Kritische Einleitung*; Eichhorn, *Einleitung in das Neue Testament*; Schleiermacher, *den sogenannten ersten Brief des Paulos an den Timotheos*; Baur, *Paulus, der Apostel*; Hilgenfeld, *Historisch-kritische Einleitung*; Holtzmann, *Lehrbuch der Distorisch-Kritischen Einleitung*.) For modern treatments of the Pauline canon especially in relation to authorship, style, and genre see van Roon, *The Authenticity of Ephesians*, 100–212; Knight, *The Pastoral Epistles*, 21–22; Porter, “Pauline Authorship and the Pastoral Epistles,” 109–10; Towner, “Pauline Theology or Pauline Tradition,” 311–14; Harding, *What Are They Saying*, 9–28; Johnson, *The First and Second Letters to Timothy*, 55–89; Fiore and Harrington, *The Pastoral Epistles*, 15–19; Aageson, “The Pastoral Epistles, Apostolic Authority,” 7–11.

² See especially Donelson, *Pseudepigraphy and Ethical Argument*, passim; Redalié, *Paul après Paul*, passim. For an alternative treatment of that same literature, see Porter and Fewster’s edited work, *Paul and Pseudepigraphy*, especially the primary sources and annotated bibliography by Baum, and the historic framing of the issue by Pitts. Baum, “Authorship and Pseudepigraphy,” 11–64; Pitts, “Style and Pseudonymity in Pauline Scholarship,” 111–15.

³ We use the term “text” here in its corpus sense—a separate textual unit within a collection of the same (a corpus), and therefore “text” is a separate NT book unless otherwise indicated. See Baker et al., *A Glossary of Corpus Linguistics*, 48; Biber et al., *Corpus Linguistics*, 12–16.

that underlying cause with little to show for it.⁴ This is all familiar ground for the New Testament scholar. So much so, in fact, that authorship debates for many are a closed issue.⁵ It is hard to see what caliber of new data—short of unearthing signed and date-stamped *autographa* perhaps—could entice NT scholars to revisit their respective stances.

The thesis of this introductory chapter is that there are actually *two* motivations to revisit this issue: compelling recent findings in *computational stylistics*,⁶ and the sheer theological implications of executing such a study on the GNT.⁷ I will present them in that order.

1.2 Recent Advances in Computational Stylistics: Its Role in “Disentangling”

It is now quite clear that recent advances in computational stylistics⁸ have informed and clarified the relationship between *idiolectic* (authorship-related) and *sociolectic* (sociocultural/sociocontextual-related) sources of variation in general and between its major components, authorship and genre, respectively. Of course, some in our guild, upon discovering

⁴ Even its most erstwhile participants and reviewers saw the discipline as deeply flawed. Holmes, “The Analysis of Literary Style,” 328–39; Neumann, *The Authenticity of the Pauline Epistles*, 23–114; Rudman, “The State of Authorship Attribution Studies,” 352–59; Forbes, “Statistical Research on the Bible,” 185–206; Parunak, “Computers and Biblical Studies,” 1112–24; O’Donnell, “Linguistic Fingerprints,” 207–26; Delcourt, “Stylometry,” 979–86; Craig, “Stylistic Analysis and Authorship Studies,” 277–84; Tuldava, “The Development of Statistical Stylistics,” 145–48; Juola et al., “A Prototype for Authorship Attribution Studies,” 170–74.

⁵ Holtzmann’s proposal to supplement F.C. Baur’s *Hauptbriefe* with 1 Thessalonians, Philippians, and Philemon remains today the consensus view of contemporary NT scholarship. A substantially smaller minority of NT scholars hold to a traditional or near traditional perspective (e.g. Ellis, “The Pastorals and Paul,” 46; Marshall, “‘Sometimes Only Orthodox’,” 15–22; Porter, “Pauline Authorship and the Pastoral Epistles,” 107–23; Johnson, *The First and Second Letters to Timothy*, 55–102).

⁶ As an aid to the reader, the first time a word considered generally specific to computational stylistics occurs, we have identified it in *bold italics*. While concerns of space do not permit a glossary in this study, many of these terms are defined in Baker et al., *A Glossary of Corpus Linguistics*.

⁷ There are three more motivations for disentangling idiolectic from sociolectic causes of stylistic variation: hermeneutical, linguistic, and experimental design reasons. These we present in Chapter Two.

⁸ The term appears to have been first used by Sedelow and Sedelow; Sedelow et al., “Some Parameters for Computational Stylistics,” *passim*. We prefer the term “computational stylistics” over “stylometry” for three reasons: (1) to distinguish it, quite simply, from the foibles of the earlier era of stylometry, (2) due to its closer association with the more modern and less problematic forms of multivariate analyses and (3) to broaden it from the historic presumptions that it addresses idiolect alone. See also Abbassi’s outline of modern computational stylistics (which he still terms stylometry); Abbasi and Chen, “Writeprints: A Stylometric Approach,” 7:3–7:10.

that computational stylistics is actually a lineal descendant of the older *stylometry*,⁹ will immediately dismiss it on that basis alone. Doing so would be both precipitous and anachronistic. Surprising as it may seem, this renewed offspring of stylometry, shorn of its positivistic pretensions and far more heavily validated than in its prior incarnations, seems at last to have delivered both insights and tools useful for just such a task. It has done so, moreover, by generating a prodigious literature of well over 1,500 peer-reviewed articles, books, and monographs, most of that total within the last twenty years.¹⁰ Four broad summary statements can be made regarding this literature.

First, earlier generations of scholars issued some vast oversimplifications regarding stylistic variation that are simply no longer tenable. In particular, earlier stylometric studies assigned, *de facto*, virtually any significant stylistic variation to authorship alone.¹¹ Now, however, close to 100 studies in traditional *textual stylistics*¹² and over 200 overall¹³ demonstrate that not only do other sources of stylistic variation *exist* (such as genre, gender, dating, audience,

⁹ See, in particular, the following concise histories of stylometry: Delcourt, “Stylometry,” 979–81; Juola, “Author Attribution,” 237–45; Holmes and Kardos, “Who Was the Author,” 5–8.

¹⁰ For more details on the meta-analysis of the literature performed for this work, see Section 2.2.

¹¹ Examples of this are legion. Notable works in biblical studies that make this equivalence despite operating within a mixed-genre corpus, or within single texts with multi-genres, include, without limitation: Harrison, *The Problem of the Pastoral Epistles*, 84–86; Morton, “The Authorship of Greek Prose,” 224; Morton, *Literary Detection*, 165–83; Bee, “Statistical Methods in the Study of the Masoretic Text,” 622; Bee, “A Statistical Study of the Sinai Pericope,” 421; Neumann, *The Authenticity of the Pauline Epistles*, 206–22. Also included here is the earlier (but not the later) Radday because this assiduous and creative scholar linguistically, experimentally and causally broadened during his career. Radday, “Isaiah and the Computer,” 73; Radday et al., “The Book of Judges Examined,” 494–99. See Chapter Three for deeper treatments of all of these scholars.

¹² By textual stylistics I mean stylistics operating upon written or spoken *language-in-use* texts. These include the various corpora studied by corpus linguists (e.g. the British National Corpus, the Cambridge International Corpus, etc.) or texts collected for historical or other reasons. The GNT is a corpus of this type. For a list of modern text corpora, see Baker et al., *A Glossary of Corpus Linguistics*, 2–6. The studies referenced are per the author’s bibliographical database.

¹³ This is arrived at by adding to that number studies from *extra-textual stylistics*. This includes idiolectic and sociolectic studies on nontraditional forms of language-in-use such as web content, computer source code, social media sources (e.g. Twitter), etc. The over 200 studies referenced are per the author’s bibliographical database.

and the like)¹⁴ but subcategories within them can be empirically distinguished as well.¹⁵

Intriguingly—and here we will return shortly—this is exactly the kind of thing that literary and linguistic theorists have been telling us were true of the literary and discourse structures within texts all along.

Second, not only is *generic stylistic variation* present in such texts, *mixed-genre* corpora very often demonstrate that a larger proportion of *the total summed stylistic variation is due to genre than authorship*.¹⁶ That is, not only do authorship and genre often dance together in texts, but genre not infrequently takes the lead. If this were not enough, in the last ten years in particular, researchers discovered that in some mixed-genre corpora an association or *covariance* exists between authorship and genre.¹⁷ This means, quite simply, that if a given author tends to write disproportionately or uniquely in a given genre, it is quite easy to confound authorship with

¹⁴ Per Wenham, “[What]...emerges from the study [is that] different genres of literature...have distinct styles”; Wenham, “Genesis: An Authorship Study,” 6. For genre separations in particular see Brainerd, “On the Distinction between a Novel and a Romance,” 260–68; Brainerd, “Pronouns and Genre in Shakespeare’s Drama,” 14–15.

¹⁵ Illustratively but not exhaustively, poetry is separated from prose or narrative from dialog in Herdan, *The Advanced Theory of Language as Choice and Chance*, 206–13; Merriam, “Invalidation Reappraised,” 419; Burrows, “The Statistical Analysis of Narrative Style,” 64; Biber, *Variation Across Speech and Writing*, 101–69. Similarly, male is discriminated from female in Argamon et al., “Gender, Genre, and Writing Style,” 326–42; Rustagi et al., “Learning Age and Gender,” 207–11; Cheng et al., “Author Gender Identification from Text,” 80–86. Lastly, authorship is differentiated by date in Temple, “A Multivariate Synthesis,” 69–74; Can and Patton, “Change of Writing Style with Time,” 66–77; HaCohen-Kerner et al., “Stylistic Feature Sets as Classifiers,” 852–59.

¹⁶ Typically in unsupervised exploratory decompositional data analysis *with mixed genres*, the first component of variation (which accounts for the most variation) tends to be genre rather than authorship. See Burrows, “The Statistical Analysis of Narrative Style,” 64; Forsyth et al., “Investigating the Authenticity of the Consolatio,” 383; Binongo and Smith, “A Bridge between Statistics and Literature,” 784; Juola and Baayen, “Authorship Identification by Cross-Entropy,” 63. Cf. Burrows, “The Interpretative Nexus between Analysis and Information,” 92–102. Cf. also Baayen et al., “An Experiment in Authorship Attribution,” no pages; section 3. Even less sophisticated methods demonstrate that generic stylistic variation tends to be greater than authorial stylistic variation. See O’Keefe, “Critical Remarks on Houk’s Statistical Analysis,” 424.

¹⁷ Factor analysis (FA) studies, at least when they are properly executed upon the covariance matrix rather than the variance matrix, often demonstrate high covariance between authorship and genre. By my count sixteen such computational linguistics studies have been performed upon the variance matrix (improperly) or covariance matrix in mixed genre corpora. In many such cases genre and authorship constitute the first two extracted components, respectively (see Chapter Two), and are orthogonal (uncorrelated) to one another. In others there is substantial covariance, and authorship and genre are extracted in the same component. For FA in biblical studies see especially Radday and Shore, *Genesis: An Authorship Study*; Mealand, “The Extent of the Pauline Corpus,” 73–79; Mealand, “Measuring Genre Differences in Mark,” 229.

genre in a mixed genre corpus, and vice versa. This previously unrecognized factor, needless to say, thickens the hermeneutical plot!

Third, in the last thirty years scholars have unleashed two great rivers of mathematical praxis upon principled collections of mixed-genre corpora. The first river, *unsupervised eigen-system mathematics* (mid-1980s to the present) separates stylistic variation into coherent and interpretable components (authorship, genre, gender, etc.). But it does more. Especially in its “high octane” formulation, *latent structural models*, the implicit directionality of these components can be said to *cause* the observed variation of the measures that comprise it.¹⁸ It is the second river, however, *supervised machine-learning and information retrieval mathematics*, that has seized the headlines.¹⁹ These supervised methods, once trained and validated, can now correctly classify unknowns (e.g. authors, genres, etc.) with accuracies exceeding ninety percent. In effect, certain of the more “plain vanilla”²⁰ classification tasks can, for all practical purposes, be considered solved (early 1990s to present).

Fourth, and most fundamentally, there is a belated but evidence-based recognition that this task remains, after all, a quintessentially *linguistic* one.²¹ That is, do we hope to decode the layers of sociocultural meaning within a corpus, replete with its specific genres, subgenres,

¹⁸ For the capability of component extraction to demonstrate directionality and hence causality see especially Davis, *The Logic of Causal Order*, 7–34.

¹⁹ Although our meta-analysis is still relatively incomplete for machine learning methods, it nonetheless contains eighty computational stylistics studies using machine learning and information retrieval (ML/IR) methods. If we include extra-textual studies (e.g. on web pages, XML, source-code etc.) that number increases to 127 studies. Intriguingly only two of those studies are upon the biblical texts; Putniņš et al., “Advanced Text Authorship Detection Methods,” J2–J13; Ebrahimpour et al., “Automated Authorship Attribution,” 6–11.

²⁰ An expression original to Koppel et al., “Authorship Attribution in the Wild,” 84.

²¹ Most computational stylistic research to date has been limited to traditional grammar. As far back as 1960, when computational stylistics was an application and not a discipline, Herdan called for a deeper linguistic grounding for the practice by calling linguists to a *deeper computational grounding*. See Herdan, *Type-Token Mathematics*, 3. Halliday two years later echoed that call by stating that “not merely are statistical descriptions needed for future progress in linguistic theory and method, they are of direct value in application.” Halliday, *Computational and Quantitative Studies*, 44. See also Craig, “Stylistic Analysis and Authorship Studies,” 29.

topics, and audience frameworks?²² If so, our muse must be language itself in the full bloom of not only its *strata* but the *paradigmatic* and *syntagmatic potentials* within it,²³ for this is the raw material by which these layers are encoded.

1.3 Implications: What if Similar Findings *Were* Verified in the GNT?

Given these four distinct movements in the recent computational stylistics literature, an obvious question follows. What would happen if the findings above turn out to be true of the GNT as well? First, and somewhat deconstructively, if the major proportion of variation in the GNT was also found to be due to genre rather than authorship—or even close to it—prior studies that have either (1) failed to test for genre as a competing theory or (2) failed to remove genre as a *covariate* have almost certainly confounded genre with authorship. Said less mathematically, if those prior studies have not controlled for the effect of genre, procedurally or in some other way,²⁴ their authorship conclusions may, ironically, be genre conclusions instead. Second, and more constructively, if a convincing separation between authorial variation and sociolectic causes of variation (including *genre/register*, subject matter, dialect, dating, gender, *audience*, and the like) can be achieved, such a thing would have broad and sweeping impact upon many *topoi* within New Testament scholarship. Not only would it influence the obvious suspects (i.e. the authorship of the Pastoral Epistles, the extent of the Pauline canon, pseudepigraphy, the

²² These are the kinds of things one would expect of the widely varying *Sitze im Leben* found in the NT. The texts within the New Testament are, after all, a grab bag of occasional documents, heterogeneous in purpose, date, ethnicity, and theological community (audience).

²³ Paradigmatics and syntagmatics, the traditional clines of language, seem to be the solid part of the dwindling residue of concepts bequeathed from Saussurean linguistic structuralism that are still relevant today. Cf. Halliday, *Computational and Quantitative Studies*, 43. Notice above that I added the word “potentials.” By this I mean to avoid two errors: (1) using an *ad hoc* unrepresentative sampling of a corpus on the one hand and (2) incorporating the unfinished and unfinishable endless regress of non-empirical Chomskyan competence exercises to our efforts. Rather we should mine something “mineable”—language-in-use within a finite corpus or a subset demonstrably stochastically representative of that corpus.

²⁴ See the early, prescient discussion of Allen in which he lists such confounding stylistics traits “...assuming that all other causes of difference such as genre, assumed audience, and age of the writer are either removed or shown not to affect the stylistic features being measured.” Allen, “Methods of Author Identification,” 906.

Synoptic Problem, and the like) it would also necessarily influence the current vigorous discourse in New Testament hermeneutics itself.

With such potent implications for the GNT, how might we get our arms around this issue theologically? What potential benefits may accrue to NT scholarship if, for instance, authorship can be cleanly disentangled from genre in the GNT? While a number of applications come to mind, perhaps it is most reasonable to explore the benefits of such an outcome in terms of its initial and most familiar quarry, the authorship of the Pastoral Epistles. This we consider next.

1.4 The Theological Potential of “Disentangling”: A Pastoral Example

Those of us with deep interest in such things are well aware that in the last decade an important milestone in Pauline studies passed with, oddly, little fanfare. It has now been over 200 years since F.E.D. Schleiermacher, following J.E.C. Schmidt, inveighed against the Pauline authorship of 1 Timothy.²⁵ When, six years later, J.G. Eichhorn extended Schleiermacher’s critique to the rest of the Pastoral Epistles,²⁶ one of the most storied dialectics in modern theology was launched in force. Now, two hundred years distant from Schmidt, many New Testament scholars would also classify Ephesians, 2 Thessalonians and Colossians as pseudepigraphal texts, leaving just seven remaining texts of the New Testament to the man from Tarsus. In just the last decade, however, in one of the vagaries that somehow manages to occur frequently regarding the Pauline canon, the focus of the debate has returned once more to where it all began²⁷—to the Pastorals themselves. What *is it* about these three gentle and winsome little

²⁵ J.E.C. Schmidt inaugurated the inquiry (Schmidt, *Historisch-Kritische Einleitung*) and was followed quickly by Schleiermacher (Schleiermacher, *den sogenannten ersten Brief des Paulos an den Timotheos*) who rejected the Pauline authorship of 1 Timothy on the basis of language and the difficulty of placing the text within the chronology of Acts. The first to judge against the Pauline authorship of all three epistles was J.G. Eichhorn (Eichhorn, *Einleitung in das Neue Testament*).

²⁶ F.C. Baur added force to Eichhorn’s critique by connecting the Pastoral’s warnings against false teachers to the Gnostic sects of the second century. Baur, *Die sogenannten Pastoralbriefe*, 8–59.

²⁷ Bond presents a terse recent summary against the Pauline authorship of 1 Timothy and by extension to the rest of the Pastorals (Bond, “1 Timothy 1:3:17,” 314–17), while a spirited and comprehensive affirmation in favor of

missives that has compelled different generations of scholars to furrow again such well-tilled dialectical ground? One answer, certainly, is that seldom in academic discourse has the extant historical data been so exquisitely configured as to allow two alternative theological constructions both so plausible and yet so antithetical.²⁸ The timing, too, was exquisite. Coupled with the momentum of the first quest for the Historical Jesus,²⁹ and propelled forward by what would soon emerge as the intellectual frontispiece of twentieth-century NT critical thought, *Formgeschichte* (form criticism), two intellectually coherent but mutually incompatible scholarly reconstructions of early Christianity emerged, each locating the Pauline texts in either separate eras or separate centuries³⁰ and each attending to a separate *Sitz im Leben*. Then too, the debate virtually pulsates with theological implications. Are these not indeed the last written musings of the apostle, now aged, at once triumphant having “finished the course,” and poignant, asking for both coat and parchments? Or are they instead the indispensable, but pseudepigraphal, eyewitness accounts of the Pauline party’s momentous second-century contest with the Valentinian *gnosis*? If the former, a monolithic critical view of the apostle must be sacrificed—here is a Paul not only capable of evangelistic pioneering but ecclesiastical polity! If the latter, the venerable history of religion school gains a helpful ecclesiastical triangulation to the

Pauline authorship was mounted by Luke Timothy Johnson. Johnson, *The First and Second Letters to Timothy*, 55–102.

²⁸ Time is always corrosive to the witness of history, of course, but the Pastorals seem to present a special case. At precisely those points where history could have had the courtesy to leave an extant witness to our debate, time has already visited and left instead lacunae. Concerning this Barrett perhaps pessimistically writes, “It must be acknowledged that in these fields there are so few historical data—and dates—that divergent views will always be more or less defensible”; Barrett, “A Critical and Exegetical Commentary,” 826. Examples demonstrating this are many but perhaps none tantalizes us more than Φ 46 which breaks off, incredibly, during 1 Thessalonians mere leaves before where (at least) 1 Timothy could have been placed. (Note: Paleographical evidence always seems capable of evoking contests. Some scholars, for instance, would place Φ 32 as early as Φ 46, and it *does* contain a fragment of Titus.)

²⁹ Momentum was all that remained given that Schweitzer made a shambles of the first quest by substantially demonstrating that each of the nineteenth-century reconstructions of Jesus make him over in the image of the scholar. Schweitzer, *Geschichte der Leben-Jesu-Forschung*, 141–565.

³⁰ Those affirming full pseudonymity (i.e. Schleiermacher followed by Baur) generally place this material in the second century.

monepiscopal urgings of Ignatius and a theological triangulation to the challenges of early-second-century Gnosticism.

Which witness is right? Both cannot be.

After 200 years, it may be time for a different perspective—one that provides a different voice based on a new kind of data, but motivated by the same desire for a more satisfactory resolution to the persistent issue of the Pauline canon.

The method already introduced, computational stylistics, might itself be considered to be such a perspective, and, for theologians at least, it clearly is a different voice! Perhaps best seen as the chastened and more mature instantiation of Radday's *statistical linguistics*,³¹ the question that concerns us is whether such a voice, *when harnessed under a careful and rigorous experimental design*,³² will say something truly distinctive as it brings its admittedly large body

³¹ Grayston and Herdan among biblical scholars were the first to use the term (Grayston and Herdan, "Authorship of the Pastorals," passim), then it was repeatedly invoked in articles and books by Radday. See especially Radday's book-length treatments. Radday, "Computerized Statistical Linguistics," passim; Radday et al., "The Book of Judges Examined," passim; Radday and Shore, *Genesis: An Authorship Study*, passim. Subsequent to Radday, the term has largely been abandoned. Interestingly, no other consensus term for the application of quantitative analytics to the NT has emerged, and hence this kind of work has operated under far flung aliases including: stylometry, stylometric analysis, authorship attribution, authenticity, disputed authorship, dialectometry, content analysis, literary computing, stylochronometry, literary style, scalometry, textual unity, and the like. A meta-analysis of these studies, however, has revealed a landscape littered with contradictory claims and counterclaims both within the biblical texts and elsewhere. Juola et al., "A Prototype for Authorship Attribution Studies," 170–71 and Rudman, "Non-Traditional Authorship Attribution Studies," 152–53; Rudman, "The State of Authorship Attribution Studies," 352–259. While principled quantitative work has been done upon the texts as we note elsewhere, per Juola and Rudman, much of it also seems to have been done at a gallop from horseback, picking "markers" at will and *ad hoc*, and selecting tests of inference in apparently the same way. This kind of work is neither formally linguistic nor formally statistical. Accordingly, we desire to revitalize the term "statistical linguistics" not only because no better eponym more tersely describes our approach, but also in hopes of reinvesting it in some measure with the promise it once carried. Statistical linguistics, as we seek to define it here, is *the application of quantitative methods to language measures developed from theoretical and applied linguistics as executed under formal principles of experimental design*. Statistical linguistics then will be the main tool we will use to develop answers to questions that reside within the larger domain space of computational stylistics.

³² Experimental design is a broad field of quantitative theory and application. As we define it here, experimental design is the formal approach to experimentation that attempts to unambiguously link causes to effects. It does this in two ways: (1) by experimental approaches that explore different subsets of the data (e.g. randomization, partitioning) and (2) by quantitative methods that can separate or disentangle multiple effects from each other and from their interactions. See Brown and Melamed, *Experimental Design and Analysis*, 1–6 for a general introduction to the aims of experimental design; Hartwig and Dearing, *Exploratory Data Analysis*; Bourque, *Processing Data: The Survey Example*, 54–78 for preparing and understanding data prior to developing a design; and Spector, *Research Designs* for data designs.

of new data to bear on that debate. In terms of our case in point, the Pastoral Epistles, it does so by addressing the core of the historic linguistic argument: that certain linguistic “markers” vary significantly between the “accepted Paulines” and other proposed Pauls, whoever they may be. All parties grant that such linguistic variation exists, but (and here lies the rub!) what does such variation actually *prove*?³³ Harrison famously proposed, if we might paraphrase his meaning, that the low-frequency markers in the Pastorals exist because it was language “out of its time.” Here were the writings of a second-century pseudepigraphist(s) being mixed in with what was considered to be, at least in Harrison’s time, texts from the first century.³⁴ In contrast to this view (and even before Harrison submitted his doctoral dissertation to the dons at Oxford), Parry published a radically different explanation for that same variation, replete with his own statistics. The observed differences seen, Parry proposed, could be fully explained by the substantially different *subject matter* that lay at the heart of these very personal letters.³⁵ As Parry was followed notably by Hitchcock, Simpson and repeatedly by Guthrie, that alternative, sociolectically-located explanation grew from subject matter to embrace circumstances, environment, addressees, and other cultural and contextual factors.³⁶ More recent work on the

³³ Carson notes that, “Undoubtedly there are differences of style in the Pastoral Epistles. The question is how to account for them.” See Carson and Moo, *An Introduction to the New Testament*, 558.

³⁴ Harrison’s argument was not based on *hapaxes* alone, but on vocabulary judged to be from the second century. See Harrison, *The Problem of the Pastoral Epistles*, 67–86. In so doing Harrison propounded the beginnings of a stylochronometric argument, although due to the earliness of its publication it necessarily lacked a formal inferential argument.

³⁵ Parry, *The Pastoral Epistles with Introduction*, cxxvi.

³⁶ Hitchcock innovatively argues that Paul’s style was necessarily affected by his immersion in Latin during his long imprisonment (Montgomery Hitchcock, “Latinity of the Pastorals,” 348). Simpson echoes these sentiments (Simpson, *The Pastoral Epistles: the Greek Text*, 20–21). Guthrie’s comments first found light in his 1955 Tyndale lecture (Guthrie, *The Pastoral Epistles and the Mind of Paul*, 27) and then echoed in his Introduction to the Pauline Epistles. These were then supplemented and reprised in his NT Introduction (Guthrie, *New Testament Introduction*, 584–622) and finally in his commentary on the Pastorals (Guthrie, *The Pastoral Epistles: An Introduction and Commentary*, 11–54). Statistical work has noted some of these same issues. Cf. Neumann’s prescient essay in McLean, ed., *Origins and Method* in which he states, “Is the answer to these puzzling results that there are other factors bringing about more variation...than authorship?” Neumann, “Major Variations in Pauline and Other Epistles,” 200.

Pastorals subsequent to the rise of *genre criticism* has focused more upon the term “genre” to summarize the omnibus idea that constructs as central as setting, occasion, and audience must necessarily express themselves in the text itself.³⁷ If we grant that the observed variation may arise from multiple sources,³⁸ our task then becomes—appropriately enough—a quantitative exercise in disentangling the sources of that variation. How *much* of the total observed variation in the texts of the GNT is due to authorship variation? How *much* is due to the specific form of communication adopted by the writer (genre)?³⁹ How *much* is due to adapting the core message to the needs of the theological community (occasion)? How *much* is due to other unspecified sociolinguistic or sociocultural sources of variation? For simplicity, this example will focus only on the causes of variation due to genre and authorship. To achieve this end, we propose a thought experiment of sorts. Short of traveling back in time to witness the events themselves, what if some benevolent being granted us a set of wishes that could, somehow, allow us to test the difference between various theories of authorship and genre? Such a wish list would surely include the following:

Wish List Item 1: Compare the relative strength (association) of the relationship between the various authorship or genre theories in the GNT and the language measures that may favor one theory or another. Ensure that the language measures range across a span of size or *rank* in the GNT (e.g. words, clauses, clause complexes). Use multiple measures of association to provide different perspectives on the underlying distribution. While we are at it, test the *direction* of that relationship. Does genre or authorship explain or predict the language forms in question (which one would NOT expect) or do the language forms explain or predict genre or authorship?

³⁷ Most would grant that the notion that the Pastorals seem to constitute a sort of epistolary subgenre has about it a certain *face validity*. Understood as literature, for instance, the Pastorals would be far closer to an informal *paraenesis* than a didactic epistle complex such as the *Hauptbriefe*. Moreover, the Pastorals clearly emulate, at least, a letter written to an individual rather than to a community.

³⁸ We submit that these three sources can be summarized under three heads: contextual, co-textual, and idiolectic, with further notional breakouts inside each. Regarding the former, greater differences in occasion, setting, and audience are necessarily more stridently advocated by historico-critical scholars than among conservative scholars. Kümmel is instructive here. After positing a later date (80 to 100) for the Pastorals, a setting in Asia Minor, addressees who are individuals rather than communities, and an occasion where the reaction to Gnosticism is more advanced, Kümmel, oddly, sees it “completely inconceivable” that the language differences could have “changed so decisively” except as explained by authorship (Kümmel, *Introduction to the New Testament*, 373).

³⁹ The idea that “forms” within language necessarily constrain the author/editor/redactor’s lexical choices is hardly novel. It is central to *Formgeschichte*.

Wish List Item 2: Develop some magic machine⁴⁰ to disentangle all the variation in all the language forms in the GNT, unsullied by other sources of variation. Let that machine discard the weakest explanations first (e.g. genre or authorship or their interactions).⁴¹ Once that disentangling work is done, use that machine to ask questions about the relationship between authorship and genre that we have been wanting to know all along. Such questions include: (1) Is authorship or genre “stronger,” and which contributes more to the observed language variation in the GNT? (2) Is authorship or genre more associated with low-frequency data? (3) How strong are the various authorship theories themselves as explanations of the observed variation in the GNT?

Wish List Item 3: Use another magic machine which takes into account all the information in the GNT fed into it and then visualizes the relationship between all the NT books in three dimensions. Let such a visualization accurately place the books at the correct distance from one another like a constellation in the night sky—but let this be a space we can actually fly through. The books (or in our analogy, stars) that are closest to one another in terms of the language of the GNT are closer to one another in our “sky.” Then step back and assess whether such a space qualitatively “looks” like genre, authorship or perhaps something in between.

Wish List Item 4: Given the long-standing pseudepigraphal assumption that unique or low-frequency forms within a language measure are differentially associated with authorship, measure whether such low-frequency information really *is* more associated with authorship as compared to genre. While we are at it, test the *direction* of that relationship as well. This measure must be a very different but complementary lens through which to view the relationship between language and authorship or genre compared to Wish List Item 1.

I am certain, of course, to have fooled no one by such a thinly-veiled Platonic “wish list.” Indeed, the telltale specificity of this list derives from the fact that modern quantitative praxis in computational stylistics addresses precisely these kinds of problems. The power of computational stylistics, as I choose to redefine it here, is that it measures items natural to the language (i.e. linguistics) while leveraging the substantial array of modern hypothesis testing and modeling methods (i.e. statistics, numerical analysis, and the like) to achieve these kinds of ends. With my ruse being so thoroughly uncovered, I can now be quite specific on how computational stylistics can address the items on our common wish list:

⁴⁰ The role of an experimental design, to be blunt, is not to prevent helpful machines from being used by researchers but to prevent both *deus ex machina* and *hominis ex machina*.

⁴¹ Radday was one of the first to posit interactions before authorship, genre, or other socio-cultural sources of variation. Radday et al., “Genesis, Wellhausen and the Computer,” 474.

- *Wish List Item 1* can be achieved by **cross tabulating**⁴² two or more frequencies⁴³ against one another, creating a table of data. In our case the rows will be a measure of language (e.g. lexemes) and the columns the theory of language (authorship or genre) that purport to best explain it. The strength of the relationship between the language measure being tested and a given theory is calculated using unbiased measures of association between the rows and its columns of the resulting table. There are two indicated statistics for testing the strength of this association: **Cramér's V**, and **Goodman & Kruskal's tau**.⁴⁴ Cramér's V is the most generally recognized symmetric (nondirectional) measure that adjusts for table size, and Goodman & Kruskal's tau is generally considered to be the best classical (not information-theory-based) directional measure of association for nominal data.⁴⁵ Directional measures test if one of the variables (e.g. a given theory of authorship) can be predicted from another (e.g. a given language measure such as lexemes). For historical reasons, statisticians refer to cross tabulation tables as *contingency tables*.
- *Wish List Item 2* uses a three-way contingency table (i.e. linguistic measure by genre theory by authorship theory) as analyzed using **Hierarchical Log Linear Analysis (HLLA)**. I will provide a more detailed explanation of HLLA in Chapter Seven.
- *Wish List Item 3* operates upon the same two-way contingency tables as in *Wish List Item 1*, but submits them to **multivariate data reduction (MDR)** and visualization. The MDR tools appropriate for contingency table analysis (e.g. lemmas by books) are **Correspondence Analysis (CA)** and **Multiple Correspondence Analysis (MCA)**.⁴⁶

⁴² Cross tabulating simply means taking two or more sets of items listed by frequency and making a table out of them. For instance, an automotive analyst for Ford Motor Company may want to know how many remaining cars by model are at his North American dealerships. Assume that the car models comprise the rows and the dealerships comprise the column. At the intersection of a specific row (e.g. a Ford Taurus) at a specific column (the Ford dealer in Topeka, Kansas) the analyst will find the number of Ford Tauruses at that dealership. The answer she seeks is the sum of all the columns, or alternatively, the sum of all the rows. The table itself is typically called a *cross tabulation table*.

⁴³ There are two broad categories of standard (Fisherian) statistics, one that operates upon "count" or categorical data (nonparametric statistics) and one that deals with scaled data (parametric statistics). For two excellent introductions to nonparametric data analysis see Agresti, *An Introduction to Categorical Data Analysis* and Reynolds, *Analysis of Nominal Data*. For the standard and more technical work see Bishop et al., *Discrete Multivariate Analysis*.

⁴⁴ For excellent introductions to Cramér's V and Goodman and Kruskal's tau see Reynolds, *Analysis of Nominal Data*, 45–60, and Liebetrau, *Measures of Association*, 13–30.

⁴⁵ Nominal data is a specific type of nonparametric data (the other type being ordinal) in which the differences between categories do not infer an ordered (ordinal) relationship between elements in the class, nor a scalar relationship (one of relative quantity whether ratio or interval scale). The numbers on the back of your favorite team's sports uniforms are classic examples of nominal data.

⁴⁶ CA is the standard way to visualize the complex relationship within a two-way contingency table. N-way tables (tables with more than two variables crossed against one another) are visualized using multiple correspondence analysis (MCA). For the inaugural work on CA and the remarkable extent to which it has been adopted in the French-speaking world, see Hirschfeld, "A Connection between Correlation," *passim* and Benzecri, "Elaboration statistique de données sur ordinateur," *passim* respectively. For the classic monograph on CA see Greenacre, *Theory and Applications of Correspondence Analysis*, and for a very accessible introduction to its application see Clausen, *Applied Correspondence Analysis*. Excellent separate monographs on MCA include those by Greenacre, and LaRoux. See Greenacre and Blasius, *Multiple Correspondence Analysis*; Le Roux and Rouanet, *Multiple Correspondence Analysis*.

- *Wish List Item 4* operates upon the same two-way contingency tables but uses **Information Theory (IT)** to analyze the association.⁴⁷ Information theory uses a distinctly different lens through which to view the data. Central to IT analysis is the idea that low-frequency information is considered more “marked”—it tells us something more *important* about the groupings than high-frequency information. Few constructs, of course, could be more relevant to our present concern given that low-frequency language forms (i.e. *hapax legomena* and the like) have been central to the linguistic argument for pseudepigraphy since the inaugural work of P.N. Harrison.⁴⁸ The standard IT statistic for this is the **uncertainty coefficient (UC)**.⁴⁹ UC predicts what percent of the column information (in our case the authorship or genre category) can be predicted given knowledge of the row information (in our case our various linguistic measures). In addition to UC, IT also provides a broad suite of measures which provide fundamental insights into complex data distributions—insights which are not available through classical Fisherian statistics.⁵⁰

⁴⁷ For clear and not overly technical introductions to information theory, see Pierce, *An Introduction to Information Theory*, 19–183; Krippendorff, *Information Theory*, 9–29. A more technical description is provided by Chirikjian, *Stochastic Models, Information Theory, and Lie Groups, Volume 2*, 272–80.

⁴⁸ P.N. Harrison was the first to assemble a detailed list of words found in only one or more of the Pastoral Epistles. See especially Harrison, *The Problem of the Pastoral Epistles*, 20, 137–39.

⁴⁹ UC is also called “Theil’s U.” See Theil, *Statistical Decomposition Analysis*.

⁵⁰ IT provides the biblical scholar with a rich suite of probes into the language of the GNT including (a) *how much* total information is being communicated by different language measures (Krippendorff, *Information Theory*, 23 and Preuss and Vorkauf, “The Knowledge Content of Statistical Data,” 23) b) *how efficiently* (Preuss and Vorkauf, “The Knowledge Content of Statistical Data,” 11–14 Krippendorff, *Information Theory*, 23) and (c) *how much noise is introduced* into that communication when the measure is grouped into different theories of genre and authorship (Krippendorff, *Information Theory*, 23). All else being equal, groupings with less noise may be considered as more parsimonious explanations of the language of the GNT.

1.5 Not So Fast: The Need for Theory

Given that such a seemingly defensible program is needed in NT studies, shouldn't we just perform the obligatory literature review within NT computational stylistics and move forward *post haste*?

Actually no, we shouldn't.

Some careful explanation is required here. I affirm that it is indeed appropriate to identify an existing research gap and then to address it with an equally appropriate methodology. The reason for not immediately proceeding is *not* that the design fails to close a number of such gaps. The design just listed, is, in fact, robust and comprehensive in many ways. It assays, for instance, more distinct linguistic measures than any prior research study published to date in the GNT—51,599 of them to be exact.⁵¹ Moreover, these measures are explicitly linguistic, probing multiple strata as well as dissecting an entire cline of language (its rank structure). The design is also consecutive methodologically. It starts with *univariate analysis* and concludes with several independent *multivariate* frameworks drawn from both *nonparametric* and *parametric multivariate analysis*. Lastly, it incorporates a dozen other experimental design features that collectively interoperate with one another to minimize Type I and Type II error.⁵² We will further detail more benefits of this design in Chapter Four. Despite all this, however, it suffers from a serious defect: regardless of how fulsome our justifications for selecting this design, *no generally accepted theory exists for doing what we hope to do—to disentangle authorship from genre*. Moreover, this theoretical vacuity has long characterized computational stylistics in general. In 1971 Todorov stated of stylistics, “there is no lack of observed facts; however...the

⁵¹ See Table 5.2 in Chapter Five. Each of these instances occur at least once in the GNT.

⁵² See Section 5.3 in this work.

uncertainty of the goal of this research hardly makes for a prosperous discipline.”⁵³ Rudman, in his 1997 LLC paper, after laying out a litany of problems with the authorship attribution arm of stylistics, concluded that the discipline *must* “...develop a complete and necessarily multifaceted theoretical framework.”⁵⁴ Juola has repeated this same sentiment (but via his preferred call for “best practices”) in a half dozen articles and conference papers between 2005 and 2012.⁵⁵ Finally, most recently (2012 as well) Argamon has called for a *general empirical theory of (automated nontraditional) authorship attribution*.⁵⁶ Each of these researchers along with others in the field agree on a diagnosis. *Computational stylistics is a discipline in search of a theory*.⁵⁷ All we are currently doing, essentially, is collecting “facts” apart from an interpretive framework (a theory) necessary to interpret and interrogate those same facts.⁵⁸

Given that computational stylistics lacks a clearly articulated theory, one might ask what, exactly, has it been doing for the last 100 years? Those familiar with the recent literature of computational stylistics, especially the reviews of Holmes, Forbes, Parunak, Neumann, Rudman, O’Donnell, Delcourt, Craig, Tuldava, and Juola,⁵⁹ already know that answer. For the vast preponderance of its modern (post-inferential) history,⁶⁰ stylometry/computational stylistics was

⁵³ Todorov, “The Place of Style,” 29.

⁵⁴ Rudman, “The State of Authorship Attribution Studies,” 361.

⁵⁵ See especially Juola and Baayen, “Authorship Identification by Cross-Entropy,” 64–65; Juola, “Authorship Attribution for Electronic Documents,” 119; Juola, “Author Attribution,” 233; Juola, “Large-Scale Experiments in Authorship Attribution,” 275.

⁵⁶ Argamon, “Scalability Issues in Authorship Attribution,” 95.

⁵⁷ Cf. de Beaugrande’s similar assessment of register in linguistics (de Beaugrande, “Register...in Search of a Theory,” 7–25).

⁵⁸ Poststructuralists and postmodernists have everywhere lauded the death of Theory as a good thing. The more sensitive among them, however, recognize that the “delegitimize-everything-but-my-own-position” praxis is itself a method of theorizing. See Moore and Sherwood, “Biblical Studies ‘After’ Theory: (Part3),” 218–25.

⁵⁹ Holmes, “The Analysis of Literary Style,” 328–39; Neumann, *The Authenticity of the Pauline Epistles*, 23–114; Rudman, “The State of Authorship Attribution Studies,” 352–59; Forbes, “Statistical Research on the Bible,” 185–206; Parunak, “Computers and Biblical Studies,” 1112–24; O’Donnell, “Linguistic Fingerprints,” 207–26; Delcourt, “Stylometry,” 979–86; Craig, “Stylistic Analysis and Authorship Studies,” 277–84; Tuldava, “The Development of Statistical Stylistics,” 145–48; Juola et al., “A Prototype for Authorship Attribution Studies,” 170–74.

⁶⁰ That is, its history after the inauguration of statistical inference by R.A. Fisher in 1922.

dominated by a *de facto* approach to the design of stylometric studies. In this approach, language “markers” were selected *ad hoc* by the intuition of the researcher, analyzed by some (ostensibly appropriate) flavor of mathematics, and significant findings were assigned to idiolectic or sociolectic causes in the same intuitive way. As studies multiplied, this praxis yielded a collision of contradictions on such a scale that Rudman, for instance, eventually concluded that there was “more wrong with authorship attribution studies than what was right.”⁶¹ Juola simply pronounced the entire field “a mess.”⁶²

Quite positively, in the last twenty years, spearheaded by Juola, Rudman and others,⁶³ experimental design bulwarks began to be put in place to reduce various sources of error. Accordingly, some of the more desultory effects of the first “mess” began to wane. What was less well-known, however, is that all was still not well in the computational stylistics kingdom. In fact, despite a shift away from the *ad hoc de facto* praxis to the more powerful supervised methods that now command the computational stylistics landscape, we are, as Argamon announced in 2012, in the midst of yet another “mess.”⁶⁴ This disarray, intriguingly, is characterized by a new set of problems that were generated, in part, by successes in solving the first “mess.” According to a second deep diagnostic by Rudman, something fundamental *still* seems to be missing from our efforts *other than* experimental design. Rudman gave a hint of this

⁶¹ Rudman, “The State of Authorship Attribution Studies,” 351. Cf. also Juola, “Authorship Attribution for Electronic Documents,” 121.

⁶² Juola, in particular, had already cemented his role as a main figure in cleaning up the mess by sponsoring the ALLC/ACH “*Ad hoc* Authorship Competition” in 2004. Juola, “Authorship Attribution for Electronic Documents,” 123–26. Nonetheless, by 2006 Juola could still accurately summarize the state of authorship attribution studies as follows: “Despite a century of research, statistical and computational methods for authorship attribution are neither reliable, well-regarded, widely used, or [sic] well understood,” Juola et al., “A Prototype for Authorship Attribution Studies,” 169. Later Argamon would recapitulate some of the same issues in relationship to automated authorship attribution. Cf. Argamon, “Scalability Issues in Authorship Attribution,” 95.

⁶³ Rudman, “The State of Authorship Attribution Studies,” 359–61; Juola et al., “A Prototype for Authorship Attribution Studies,” 173–76. In a follow-up article, Rudman grudgingly grants that “Fourteen years later we can see a muddled but gradual advancement in some areas.” See Rudman, “The State of Non-Traditional Authorship Attribution Studies-2012,” 259.

⁶⁴ Argamon, “Scalability Issues in Authorship Attribution,” 95.

as far back as his 2003 ALLC/ACH conference paper, in which he wrote, “Most nontraditional authorship attribution studies place too much emphasis on statistics, stylistics and the computer and not enough focus on the...text itself.”⁶⁵ Argamon, moreover, adds other problems to the latest list:

While many significant accomplishments have been achieved, the field is highly fragmented, with little to no general theory or deep understandings of the strengths and weaknesses of different methods. Indeed, there is very little agreement, if any, on standard evaluation methods, so it is nearly impossible to really measure progress in the field as a whole.... There have been few attempts at large-scale systematic experimentation...with the goal of deriving *more general understandings* than particular studies can give.⁶⁶

What we have, then, are evolutionarily more powerful mathematical methods than ever before (per Section 1.2), tools even capable of producing *point solutions* to certain classes of problems in computational stylistics,⁶⁷ but, yet again, *no general theoretic framework* to “measure progress in the field as a whole.”

⁶⁵ Rudman, “On Determining a Valid Text for Non-Traditional Authorship Attribution Studies: Editing, Unediting, and De-Editing,” 18.

⁶⁶ Argamon, “Scalability Issues in Authorship Attribution,” 95. (Italics mine)

⁶⁷ See Section 2.4.3.

1.6 The Central Objective and Structure of This Work

In agreement with the separate assessments of Juola, Argamon and Rudman, a general empirical theory is needed—but not just for authorship attribution. A general theory is needed for the whole of computational stylistics, one that embraces the three discrete tasks within it: (1) *classification* (e.g. distinguishing different authors, genres, and the like), (2) *decomposition* or disentangling stylistic variation into its components and (3) *causation* or identifying which observed or latent (unobserved) components feed into other components that, in turn, help us arrive at a cause.⁶⁸

We seek, then, to achieve three separate objectives in this work: (1) to provide an initial working proposal for a theory of disentangling genre from authorship, (2) articulate an experimental design for such a theory and (3) produce initial data toward that end. Given these objectives we are now free to clarify the structure we will adopt in the remainder of this work. Four elements are central to that structure: ground-clearing, history, method and praxis.

Ground-Clearing. A work this interdisciplinary requires more than the standard amount of ground-clearing. Accordingly, in Chapter Two four separate ground-clearing exercises will be explored. *First*, the central methodological approach of this work, *abduction*, will be defined and justified in terms of the Greek New Testament. *Second*, the large gaps (or desiderata) that exist in NT computational stylistics will be detailed along with their implications. *Third*, I will provide two primers, one on computational stylistics and one on linguistics, to establish a framework for our task. *Fourth*, initial components of that framework will be proposed and interrogated as to their adequacy in constituting the first cycle in a multi-cycle abductive design.

⁶⁸ Moreover, general experimental design theory teaches us that successful theories are often *hierarchically related*. That is, once the principles for a general theory are secured, many of those same principles may be inherited as a class for the more specific problems. Regardless of whether such a thing proves to be true in this case, we need to keep our eyes open for generalized solutions.

History. In Chapter Three we will review the literature. But we propose a modest variant of the same, an *experimental design-based review* of the literature, one that selects from over 1,500 reviews, articles, books, and monographs those that best typify the failures and successes of the respective eras in stylometry/computational stylistics. Our goal will be to draw from these studies a set of *best practice* principles for disentangling authorship from genre.

Method. In Chapter Four we will combine the retrospective experimental design review from Chapter Three with experimental design theory to develop an overall abductive theory and experimental design for disentangling authorship from genre. This will yield a synthetic method for the computational analysis of the GNT. While the outcome of this work is essentially a “measure by method” matrix of subexperiments, it is also a matrix that is abductive and reentrant. That is, the results of certain experiments can change the presence or priority of existing experiments in the design matrix.

Praxis. In Chapters Five through Nine I will execute the first cycle through the experimental design articulated in Chapter Four. The data from these five chapters will be analyzed and integrated to draw the conclusions found in Chapter Ten.

In my view, this broad approach—history, method, and praxis—is the central path needed to overcome prior design problems and bring new data to bear on disentangling genre from authorship in the Greek New Testament.

2 Ground Clearing

2.1 Developing a Theory of Disentangling: Proposing an Abductive Approach

The goal of disentangling genre from authorship will require some sort of organizing framework. Because we seek to make progress toward not only a single experimental design but one that evolves, I propose an *abductive* experimental framework going forward. Abduction is, by now, a well-recognized fixture on the landscape of both epistemology and applied experimental design. For a more formal introduction to its mathematics, its emergent role in post-structural epistemology, and its utility in postpositivist experimental designs, see Peirce,¹ Reichertz, Hoffman, and McKaughan.² Some may challenge the appropriateness of abduction for such a task, given its very meager application in literary theory³ and its complete absence in NT computational stylistics.⁴ To this objection it might be noted that this paucity may be more related to the lack of experimental designs of any stripe in these traditionally diachronically-dominated disciplines.⁵ More relevantly perhaps, it has become increasingly apparent that the actual praxis of theory-making and experimentation is a thoroughly abductive one.⁶ Accordingly, while its mathematical basis and the full explication of its intersection with traditional

¹ See especially, Peirce, "On the Natural Classification of Arguments," 261–87; Peirce, "Deduction, Induction, and Hypothesis," 470–82; Peirce, "On the Logic of Drawing History from Ancient Documents," 75–114. Concerning the likely unrecognized connection between Peirce and Polanyi, see Mullins, "Peirce's Abduction," 199–216.

² For the abductive turn in research, and its mediating position between induction and deduction, see Reichertz's helpful introduction, Reichertz, "Induction, Deduction, Abduction," 124–30. For delimiting but not debilitating critiques of the role of creativity/methods of inquiry in Peirce's abduction, see Hoffmann, "Problems with Peirce's Concept of Abduction," 271–305. Cf. also Paavola, "Abduction as a Logic and Methodology of Discovery," 267–81. For the various interpretations of abduction, see McKaughan, "C. S. Peirce, Abduction, and the Pursuit of Scientific Theories," 446–60.

³ Literary abduction is still largely exploratory (Bunt and Black, *Abduction, Belief, and Context*; McGann, "Marking Texts of Many Dimensions," 213).

⁴ For theological arguments using Peirce's logic, however, see Sisson, "Abductive Logic and Rhetorical Structure," 98; Ochs, "Philosophic Warrants for Scriptural Reasoning," 465–79.

⁵ One mediating approach that lies between the enlightenment era objectivism and the post-Gadamerian philosophical hermeneutic can be found in mixed methodology methods of verification. Cf. Van Gelder, "Method in Light of Scriptures," 62.

⁶ See especially Montgomery, "Theologian's Craft," 73.

experimental design theory is beyond the scope of this work, two characteristics of abduction are relevant and quite important here: its philosophical grounding and its pragmatic utility.

Philosophically, abduction borrows from the classical strengths of deduction (syllogistic reasoning) and induction (design and execution of experiments). In brief, it seeks to avoid their respective weakness. This is not limited, of course, to faulty premises in the case of deduction, and philosophical naiveté in terms of induction. One of the most terse and useful descriptions of abduction for theory development is provided by D. A. Carson (and in a theological context no less):

...any complex theory in virtually any field (some branches of mathematics might be excluded) is built up by a mixture of deduction and induction—indeed, by more than these two, but by what is variously called retroduction, abduction or adduction. The method is found in the natural sciences as well as in the humanities and constitutes a major part of theory formation and justification. In the views of most theorists, adduction (as I shall call it) is not so much separate from deduction and induction as a category that retains both while going beyond them to describe the creative thought, the sudden links and the establishment of paradigms that account for the evidence as accumulated and understood to that point.⁷

While Peirce's earlier writings were somewhat unclear concerning the relationship of abduction to both induction and deduction, an oversight he would later admit,⁸ abduction can be thought of as the *poiemic*⁹ engine behind a fulsome, reentrant, experimental design, praxis, and evaluation loop. This is a loop in which the researcher is free to exchange, integrate, or discharge any given set of priors, procedures or practices in the search for a more useful or adequate “outcome,”—however that outcome may be defined.¹⁰ For this reason, abduction is often termed

⁷ Cf. Carson, “Three More Books on the Bible,” 343.

⁸ Pierce admits this in a letter to Paul Carus circa 1910; Peirce, *Collected Papers of Charles Sanders Peirce*, Paragraphs 227–28.

⁹ I owe this terminology to Bruggemann who issues a call to “fund” (i.e. provide the raw material for) a new world of interpretation, one that “counter-imagines” the world. Bruggemann, *Texts Under Negotiation*, 20. Bruggemann's vision itself was animated in part by Richard Kearney's call to question postmodernity's impress as the final port of call for philosophy. Kearney calls for a reimagining that is critical but one that also embraces ποιησις, the possibility of optimistic, creative work. See Kearney, “Ethics and the Postmodern Imagination,” 43–45.

¹⁰ Here “fitness” in computational stylistics is to be understood as the *goodness of fit* between the projections yielded by the models used and their actual distributions as found in the GNT.

inference to the best explanation, or IBE. I see the relationship of abduction and IBE, however, on a cline. Abduction is the *poiemic* engine (the justification for the in-breaking of creativity from outside the system) and demands a process, with IBE being that process.¹¹ For simplicity we will ignore this distinction and refer to both as abduction.

Pragmatically, the utility of abduction for our purposes can be located in two aspects; first, its view of inference and second, its appropriateness as a framework for research. We submit that individual instances of abductive inference are of little aid, but that abduction, when employed within a persistent, repetitive design loop—one characterized by multiple intersecting methods of verification—can make progress if indeed the core problem is tractable. An example may suffice to demonstrate the relationship between *abductive inference* and *abduction as an iterative process*. Peirce's complex symbolic mathematics on abductive inference can be reduced without much loss of specificity to the following notion expressed syllogistically:¹²

P1: The surprising fact, C, is observed.

P2: But if H [an explanatory hypothesis] were true, C would be a matter of course.

P3: Hence, there is reason to suspect that H is true.

Replacing the symbols with one such relevant NT construct would yield:

P1: The surprising fact that the Pastoral Epistles display a disproportionately high proportion of *hapax legomena* is observed.

P2: But if the Pastorals reflect Pauline authorship communicated through a distinctly different genre/register, audience, and occasion from the rest of the GNT (and by definition subgenres select from infrequent register-specific lexical and grammatical categories) *hapax legomena* would be a matter of course.

P3: Hence, there is reason to suspect that the Pastorals reflect Pauline authorship.

In order not to prejudice the argument unduly, however, another such construction would also follow:

¹¹ Two recent important articles in IBE/abduction should be consulted. The first discusses its metaphysical implications (per Kripke, "Outline of a Theory of Truth," 699-716) while the second relates to its physical/scientific theory-making utility (Maddalena, "Abduction and Metaphysical Realism," 243-59; Magnani, "An Abductive Theory," 261-86). For the non-identity of abduction and IBE, see Mackonis, "Inference to the Best Explanation," 976-78.

¹² See Paavola, "Abduction as a Logic and Methodology of Discovery," 267-68.

P1: The surprising fact that the Pastoral Epistles display a disproportionately high proportion of *hapax legomena* is observed.

P2: But if the Pastorals are written by a second-century pseudepigraphist whose second-century vocabulary flooded into the text quite naturally, then accordingly, a high proportion of *hapax legomena* would be a matter of course.

P3: Hence, there is reason to suspect a second-century pseudepigraphist wrote some or all of the Pastoral Epistles.

Seen in this way we apparently have arrived only at a more syllogistically sophisticated stalemate than we knew before. We still do not know if we have a Paul (per Section 1.4) capable of evangelistic pioneering as well as ecclesiastical polity! A single instance of abductive inference, then, seems to be of little help. When expanded to a *theory of experimentation*, however, the utility of abduction is more apparent. What might happen, for instance, if the dual P1 statements above were not limited to *hapax* only? What if the italicized words in both P1 statements above were replaced by thousands, or perhaps millions, of the separate stratal, paradigmatic, syntagmatic and functional linguistic measures found in *greater or lesser* relative normed frequency in the Pastorals than, say, in the rest of the GNT?¹³ Assume, then, that we built those inferences and executed all those experiments. The questions would then follow: “For those measures where statistically significant differences are found (either individually, or as a coherent class) what cause better explains which findings? Idiolectic variation? Sociolectic variation? Dialectic variation? And how would we determine such a thing?” These are the kinds of things an abductive design can address. This, in turn, begs another question: pragmatically, what would such a reentrant, abductive theory in the GNT look like? Inspired by the creative abduction notion discussed by Prendinger and Ishizuka,¹⁴ I propose that an adequate abductive

¹³ In the research in this work, for instance, which operates only at the syntagmatic cline of the Koine, we have found 51,599 linguistic measures that occur at least once in the GNT. If we assume that one-tenth of those exist in high enough frequency to be tested with, say, classical Fisherian inference, this yields over 5,000 such inferences.

¹⁴ Prendinger and Ishizuka, “A Creative Abduction Approach,” 324–26.

framework for theory development and praxis addressing these concerns circumscribes a loop of seven steps:

1. *Problem Definition*: articulating the main problem to be solved
2. *Methodology Articulation and Definitions*: defining starting stances (priors), terms, secondary problems (if any), and needed tools
3. *Mixed Methodology Data Collection*: collecting relevant qualitative as well as quantitative data to inform the experimental design
4. *Experimental Design Articulation*: developing an experimental design that incorporates the present state of items 1-3
5. *Experimental Design Execution*: performing a subset of the experiments of the design
6. *Experimental Design Evaluation/Conclusions* including abductive improvements suggested by any findings obtained
7. Reentry to the process by returning to step 1

Table 2.1 below lays out the more complete Prendinger and Ishizuka-inspired design, and the sequence to be followed in this work.¹⁵

<i>The Modified Prendinger and Ishizuka Abductive Framework Followed in This Work</i>				
<i>Step</i>	<i>Phase</i>	<i>Tasks</i>	<i>Section</i>	<i>Section Title</i>
1	Problem Definition	Define the Primary Problem to be Solved	1.1-1.6	The Main Problem: We Lack a Theory for Disentangling Authorship from Genre
2	Methodology Articulation and Definitions	Propose, Justify Approach	2.1	Developing a Theory of Disentangling by Proposing an Abductive Approach
		Define Priors	2.2	Articulating Starting Stances (Priors)
		Define Terms, Criteria		
		Define Any Secondary Problems to be Solved	2.3	Gaps in the Current State of NT Computational Stylistics
		Define Needed Tools	2.4	Computational Stylistics: A Primer
		Define Existing Solutions	2.5	Linguistics: A Primer
			2.6	Before Beginning: Did We Miss Anything?
3	Mixed Methodology Data Collection	Collect the Relevant Qualitative Data	3.1-3.17	An Experimental Design Review of the History of the Discipline
		Collect the Relevant Quantitative Data	3.18	Summarizing the Experimental Design Review
4	Design Articulation	Build Out the Experimental Design	4.1-3	Implications from Prior and Current Work
			4.4-6	Developing the Synthetic Experimental Design
5	Experimental Design Execution	Defining Testable Language Measures, Authorship Theories and Genre Theories	5.1	Framing the First Cycle in the Design
			5.2	Developing Testable Language Measures
			5.3	Developing Testable Authorship Theories
			5.4	Developing Testable Genre Theories
		Results Part I: Testing GNT Authorship and Genre Using Univariate NPA and IT	6.1	Preparing the GNT for Analysis
			6.2	Interpreting Contingency Tables: A Primer
			6.3	Testing A/G via Nonparametric Association
			6.4	Testing A/G via Information Theory Measures
		Results Part II: Disentangling GNT Authorship and Genre Using Multivariate NPA	7.1	Disentangling Using HLLA: A Primer
			7.2	H1: Is Authorship or Genre “Stronger”?
			7.3	H2: Is A/G Associated with Low Frequencies?
			7.4	H3: How Do the Strongest Theories Compare?
			7.5	Disruptive Abductive Finding on Authorship
		Results Part III: Visualization of GNT Texts Using EMVA	8.1	Visualizing GNT Texts: A Primer
			8.2	Experimental Design (CA)
			8.3	Interpreting CA
			8.4	The CA Data Itself
			8.5	GNT Text Visualization via MCA
		Results Part IV: How Abduction Makes Us Follow Bread Crumbs: Exploring GNT Information Loss Via Univariate IT	9.1	The History of IT in Linguistics
	9.2	Information Theory: A Primer		
	9.3	Q1: Which Contains the Most Information?		
	9.4	Q2: Which Encodes the Most Information?		
	9.5	Q3: How Does TI Loss Differ by Lang. Rank?		
	9.6	Conclusions		
6	Conclusions and Next Steps	Conclude and Develop Changes to the Experimental Design	10.1	Finding
			10.2	Implications for NT Studies
			10.3	Implications for NT Computational Stylistics
			10.4	Directions
			10.5	Hermeneutical Reflection

Table 2.1

¹⁵ Notice that the design first underscores the main problem; we lack a theory of “disentangling.” The gaps and contradictions in GNT are the secondary yield of that primary problem. Accordingly, they are relegated to step 2.

Some clarifications regarding the terminology in Table 2.1 are appropriate here. First, Table 2.1 is the overall framework (the tasks) that will be followed within this dissertation. The overall experimental design, the “blueprint” for the multi-year research project that this dissertation inaugurates, in contrast, is to be found at the end of Chapter Four. Second, because most NT scholars do not define themselves as either linguists or computational stylists I have included primers on those subjects in this dissertation. In addition, other quantitative primers will be interspersed where they are actually used, primarily in Chapters Five through Nine.

2.2 Articulating Our Starting Stances (Priors)

One of the benefits of abduction is that one can start where one chooses, with any set of presuppositions or priors. This is because the process of abduction is both verifiable and reentrant. In theory, therefore, after several iterations the researcher will move toward a better set of priors, methods, and experiments that better fit the qualitative and quantitative data.

Granting the above, however, our goal is to begin with the most robust set of starting priors and methodologies possible, to reduce our “cycles” through the abductive loop. A second characteristic of this starting set is that it may be as large or small as the researcher desires. While no such exercise can fully escape the charge of being labelled idiosyncratic, making this charge quite misses the abductive point. No longer can privileged foundational priors deprivilege our choice of a starting set, regardless of how large that set may be. The important thing here is to *begin* with as relevant a set of those starting methods as possible and to articulate and defend those priors. Accordingly, we are now prepared (and required per section 2.2) to state the initial starting position for our work. Based on a review of the literature in Chapter Three, and a limited set of other priors the author has chosen to defend, five starting stances can now be articulated. Those stances and their justifications are as follows:

Philosophically and Hermeneutically: (1) I accept Tyson’s critique of naïve empiricism and oppose the still unfortunate ubiquity of praxis in which researchers neither state their priors, nor articulate the domain limitations of their conclusions.¹⁶ (2) I reject the equally naïve perspective that erroneously equates positivism and empiricism (e.g. Spencer’s positivist-empiricism).¹⁷ (3) I affirm Argamon and Rudman’s call to triangulate findings in computational stylistics with more traditional (qualitative) approaches.¹⁸ (4) I affirm Tashakkori and Teddlie’s critique of positivist-only verification and affirm the utility of post-positivistic (*mixed method*) verificational schemata in data evaluation.¹⁹ (5) I affirm the parallel use of mixed methodology verification alongside Bayesian and Fisherian verification. (6) I affirm Kripke’s *principle of contingent necessary truths*.²⁰ To not do so would deprivilege potentially relevant data. (7) I reject abductive vacuity, that is, the praxis of rigid, non-reentrant experimental designs. Polanyi, Peirce, Kripke, Mullins, Paavola, and Tashakkori are important conversation partners here among many others.²¹

Linguistically: (1) I affirm that since the raw data is quintessentially linguistic, the measures used should be thoroughly linguistically informed. (2) I recognize that multiple linguistic frameworks exist. I propose, however, that for both theoretical and pragmatic reasons (see Section 2.5) *Systemic Functional Linguistics* (SFL) presents the most adequate framework

¹⁶ Tyson, “Transcendence and Epistemology,” 256–62.

¹⁷ This refers to the mistaken notion that the philosophy of empiricism is necessarily positivism—hence the hyphen. While Spencer’s lockstep apposition of these separate ideas is anachronistic today, it certainly rang true in terms of mid- to late-twentieth-century stylometry. See Spencer, “Preparing the Way of the Lord,” 109. Cf. also Morgan who accurately critiques the frequent postmodern educator’s view of science as monolithic. Morgan, “Poststructuralism and Applied Linguistics,” 1035. All such oversimplifications are suitable only as a strawman.

¹⁸ Rudman terms this a failure to look at “all of the external and traditional evidence.” Rudman, “The State of Authorship Attribution Studies,” 359. Cf. especially how Argamon uses systemic functional theory to develop feature sets for computational analysis; Argamon and Koppel, “Finding Meaning in Stylistic Variation,” 88–89.

¹⁹ Tashakkori and Teddlie, *Handbook of Mixed Methods*, 4–17.

²⁰ Kripke, “Outline of a Theory of Truth,” *passim*.

²¹ Polanyi, *Science, Faith, and Society*, 21–41. For the other conversation partners, see Section 2.1.

for developing those linguistic measures.²² (3) I affirm that comprehensive SFL measures should be developed, inclusive of the multiple strata, syntagmatic structures, paradigmatic systems, and metafunctions of the GNT. (4) I affirm that *traditional grammar*²³ is of questionable value in such an exercise except as a control. The reason for this is that the “forms” are completely subsumed within the far more numerous (granular) functional categories of SFL. Among other theorists, Halliday, Martin,²⁴ Fawcett,²⁵ and Bell²⁶ are important here.

Quantitatively: (1) I recognize that *unsupervised covariance structure* research of texts which began to be explored via *unsupervised multivariate techniques* in the late 1980s has been largely overshadowed by *supervised multivariate techniques* and their outcomes. (2) I recognize that unsupervised techniques, however, because they retain the entire variance and covariance information of the corpus in question are best at elucidating the structure of the relationship between idiolect and sociolect. (3) We recognize that while supervised techniques are efficient at creating point solutions, those solutions tend to be overfit and not generalizable. (4) We recognize that overfitting can be minimized by various methods of cross/hold-back-sample validation. (5) We deny that the various methods of validation make supervised solutions truly *generalizable*. This is because supervised techniques, at least as they are traditionally executed, obliterate the structural relationship between the various fractions of stylistic variation (e.g.

²² For a selected bibliography of Halliday’s theory and its application to computational stylistics see Halliday, *Halliday: System and Function in Language*; Halliday, *Explorations in the Functions of Language*; Halliday et al., eds., *New Developments in Systemic Linguistics*; Halliday, *On Grammar*; Halliday, *Linguistic Studies of Text and Discourse*; Halliday, *On Language and Linguistics*; Halliday and Matthiessen, *An Introduction to Functional Grammar*; Halliday, *Computational and Quantitative Studies*.

²³ By comprehensive we mean research that either uses a token vector of all types (i.e. discrete categories) for a given system, structure, strata or metafunction, or feature set selection that identifies a representative subset of the same.

²⁴ See Martin, *English Text*; Martin and White, *The Language of Evaluation*; Martin and Rose, *Genre Relations*.

²⁵ See Fawcett, *Cognitive Linguistics and Social Interaction*; Halliday et al., eds., *New Developments in Systemic Linguistics*; Fawcett, *A Theory of Syntax*.

²⁶ See especially Bell, “Language Style as Audience Design,” 145–204.

between sociolectic variation and idiolectic variation) by using only those linguistic features that best classify known texts using that prior information. (6) I recognize that hermeneutically rich layers of textual variation (genre, dating, authorship, and the like) have been modestly excavated by a modest number of researchers using eigen-systems mathematics. It should be underscored, however, that such excavation has been largely limited to the linguistic rank of lexis (spans of linguistic structure no longer than one word long). (7) I affirm on the basis of the five prior observations, that both latent structural analysis via Structural Equation Modeling (SEM)²⁷ and graphical models²⁸ are clearly indicated. Long, Pearl, Burrows, Biber, and Jordan,²⁹ are pioneers in this area outside of biblical studies.

Experimentally: (1) I recognize the central importance of experimental design in the elimination of Type I and Type II error. (2) I affirm that a broad-based initial exploratory phase of experimentation is called for, requiring the integration of both supervised and unsupervised knowledge discovery and classification in texts. (3) I affirm that *a posteriori* quantitative methods of extracting authorship and genre factors “up from the text” (e.g. via eigen-system mathematics) should be compared to *a priori* quantitative methods of testing. These are, quite relevantly, the accepted categories of authorship and genre which come “down from

²⁷ Classic latent covariance structure modeling (SEM, or LISREL) has been only rarely deployed in computational stylistics. Some tentative studies have been pursued but these have not been executed on the dual clines of language. See Gries, Biber, and Paolillo et al. (Gries, *Multifactorial Analysis in Corpus Linguistics*, 178–80; Biber, *Dimensions of Register Variation*, 140; Biber, *Variation Across Speech and Writing*, 115–20; Biber, “Linguistic Analyses of Genre Variation,” 340; Paolillo et al., “Genre Emergence in Amateur Flash,” 277–301). For nonparametric latent structure work, see Dai and Storkey, “The Supervised Hierarchical Dirichlet Process,” *passim*. In contrast, latent models have been widely used in document and topic classification (e.g. Savoy, “Authorship Attribution Based on Specific Vocabulary,” 347–48; Jockers, *Text Analysis with R*, 135–39), yet again, not upon the biblical texts.

²⁸ Structural models and graphical models recognize that in complex systems latent, unobserved variables exist between observed effects and a presumptive cause. Graphical models such as Bayesian networks are often used in topic and genre classification.

²⁹ See especially the following: Long, *Confirmatory Factor Analysis*, 11–81; Long, *Covariance Structure Models*, 19–85; Pearl, *Probabilistic Reasoning in Intelligent Systems*, 1–19, 29–73; Burrows, “The Interpretative Nexus between Analysis and Information,” 91–103; Biber, “Linguistic Analyses of Genre Variation,” 332–44; Biber et al., *Corpus Linguistics*, 84–242; Jordan, “Graphical Models,” 141–54.

theory/history” of NT scholarship (e.g. via log-linear modeling). (4) I affirm the necessity of articulating a formal experimental design theory that achieves these ends inclusive of experiments and sub-experiments. Fisher, Spector, Brown and Melamed, Gray, Atkinson, and Meyers are helpful here.³⁰

2.3 Gaps in the Current State of NT Computational Stylistics

2.3.1 Gaps in Our Current Knowledge Compared to Those Starting Stances

Now that we have articulated our starting stances above, our next task is to identify what is missing in regard to disentangling genre from authorship in the GNT. One way to do so, and admittedly the most coarse, is to query our meta-analysis. That query reveals that out of a literature of 235 relevant multivariate primary studies in idiolectically or sociolectically-related causes of stylistic variation published since the mid-1980s, studies of idiolect outnumber studies of sociolect almost three to one. Moreover, only twenty studies of any multivariate stripe have been performed upon the Greek GNT. This relative paucity of research in NT computational stylistics becomes clearer when we compare it against the five-fold starting framework just articulated in Section 2.2. That analysis reveals the following gaps in regard to the GNT.

Philosophically and hermeneutically, no computational stylistics studies in the NT have articulated an approach that addresses the seven central philosophical and hermeneutical objections raised by (1) Tyson's critique of naive empiricism, (2) Spencer's positivist-empiricism, (3) Argamon and Rudman's separate requirements to triangulate findings with more traditional (qualitative) approaches to authorship or style (in our case the long history of accepted diachronic praxis in NT studies, (4) Tashakkori, and Teddlie's critique of positivist-only

³⁰ See especially the following sections of these important works: Fisher, *The Design of Experiments*, passim; Spector, *Research Designs*, 39–76; Brown and Melamed, *Experimental Design and Analysis*, 20–44, 50–79; Gray, *Doing Research in the Real World*, 15–62; Atkinson, *Optimum Experimental Designs, with SAS*, 25–33, 58–71; Meyers et al., *Applied Multivariate Research*; Rasch et al., *Optimal Experimental Design with R*, 175–288.

verification, (5) the lack of integration of Bayesian and classical inference with mixed methods approaches, (6) Kripke's principle of contingent necessary truths, and (7) the exercise of an abductive method—that is, the avoidance of relevant data on the one hand and the praxis of rigid, non-reentrant experimental designs on the other. In my reading no single published work to date has addressed any more than two of these issues. Watson, Thiselton, Bockmeuhl, Luz, Aichele, and Moore are important conversation partners here for NT scholars.³¹

Linguistically, no studies have attempted to address a relevant problem in NT studies by using a *linguistically* comprehensive set of measures for either the syntagmatic structures, paradigmatic systems, multiple strata or multiple metafunctions of the GNT.³² This is quite crucial because if we presume, for instance, that the systems of the Koine are at least as delicate (detailed) as they are in English, NT scholars have likely empirically studied somewhat less than one percent of the linguistic potential in the GNT.³³ Among NT linguists, Silva, Nida, Pitts, O'Donnell, and Porter are important here.³⁴

Quantitatively, no studies of the covariance structure of the GNT across the systems and rank structure of the Koine have been performed. Moreover, while a handful of studies of the hermeneutically rich layers of NT textual variation (principally in terms of genre and authorship)

³¹ See especially Watson, *Text and Truth*, 26; Bockmeuhl, *Seeing the Word*, 30–39; Luz, *Studies in Matthew*, 313–32; Aichele et al., “Historical-Critical and Postmodern Interpretations of the Bible,” 383; Moore and Sherwood, “Biblical Studies ‘After’ Theory: (Part 2),” 90; Thiselton and Marko, “Thirty Years of Hermeneutics,” passim.

³² By comprehensive we mean research that either uses a token vector of all types for a given system, structure, strata or metafunction, or feature set selection that identifies a representative subset of the same.

³³ That is, the discrete linguistic measures and their combinations in the multiple strata, paradigmatic systems, syntagmatic structure and multiple metafunctions of the Koine contemporary to, and including the texts of the GNT. See section 2.6 for a justification of this percentage.

³⁴ See especially, Silva, *Biblical Words and Their Meaning*, 101–78; Nida and Louw, *Lexical Semantics of the Greek New Testament*, 107–32; Porter and Pitts, “New Testament Greek Language and Linguistics in Recent Research,” 215–36; Porter and Pitts, eds., *The Language of the New Testament*, passim; O'Donnell, *Corpus Linguistics*, 273–485; Porter, *Verbal Aspect in the Greek of the New Testament*; Porter and Carson, eds., *Biblical Greek Language and Linguistics*; Porter and Carson, eds., *Discourse Analysis and Other Topics*; Porter, *Studies in the Greek New Testament*; Porter and Reed, *Discourse Analysis and the New Testament*; Porter and Carson, eds., *Linguistics and the New Testament*.

have been excavated by Merriam, Mealand and others³⁵ using eigen-systems mathematics,³⁶ that excavation has been disproportionately limited to syntagms no longer than a single word.

Therefore, more unsupervised decompositional work at all layers of language is indicated.

Lastly, given that we know virtually nothing about the structures of variation and covariation in the GNT, Structural Equation Modeling (SEM) and graphical models are clearly indicated.

Experimentally, no studies in the GNT have compared and contrasted supervised and unsupervised knowledge discovery and classification in the GNT. Nor have unsupervised *a posteriori* quantitative methods of extracting authorship and genre factors “up from the text” (e.g. via eigen-system mathematics) been systematically compared to *a priori* quantitative methods. (The latter relate to categories of authorship and genre which have come down to us from the history NT scholarship). Lastly, no formal experimental design has been articulated that achieves these ends with the experiments and subexperiments necessary to reduce Type I and Type II error.

2.3.2 Gaps in Sociolectic *Desiderata*

Given the relative disproportion of idiolectically-related to sociolectically-related studies discussed above, and the virtual absence of NT sociolectic studies, it is not surprising that many sociolectic gaps exist in the GNT. The questions below summarize a sampling of those sociolectic gaps.

- How many layers of sociolectic stylistic variation exist in the GNT (e.g. genres, subgenres, stylochro-metric, topical, gender, ethnic, audience, etc.)?
- Do those layers map cleanly to identifiable causes (authorship, genre, audience, etc.)?
- How much of the total stylistic variance is accounted for by each layer? How does that compare and contrast to idiolectic layers?

³⁵ See especially Radday and Shore, *Genesis: An Authorship Study*; Neumann, *The Authenticity of the Pauline Epistles*; Ledger, “An Exploration of Differences in the Pauline Epistles,” *passim*; Mealand, “The Extent of the Pauline Corpus,” *passim*; Mealand, “Style, Genre, and Authorship in Acts,” *passim*; Putniņš et al., “Advanced Text Authorship Detection Methods,” *passim*; Ebrahimpour et al., “Automated Authorship Attribution,” *passim*.

³⁶ Largely factor analysis, principal components analysis, and discriminant analysis.

- How correlated are these layers? Do they stand alone (display only “main effects”) or are there interactions between the layers?
- If there are interactions, how deep does the “rabbit-hole” go in these interactions? One-way? Two-way? Three-way?
- Which paradigmatic systems and syntagmatic structures contribute disproportionately to which “layers” in the text?
- When an association is observed between a certain linguistic unit (say, clauses) and a given sociolectic layer (say, genre) is that association directional? If so, is it causal?
- What proposals of the functional linguistics schools (Copenhagen? London?) are most fruitful in serving up those kinds of associations?
- What causal structure yields the highest fit (coefficient of validation) in explaining the final textual structure³⁷ of a given corpus such as the GNT?
- Which tokens among the types (e.g. the elements among how we class them) contribute most to the conclusions arrived at in the previous questions?
- How would the questions above change if we enlarge our scope to include the wider context of the entire corpus of the Koine contemporaneous to the GNT?
- How do the answers above change when we increase the radius of our investigation in the Koine from person to family to clan to village/tribe within a language group?
- How do all these answers change if the audience is singular or plural?
- How do all these answers change if the audience differs greatly in social status (e.g. *mandata principis* literature) or *power-distance* from the writer?³⁸

Given the above list, it is clear that despite the recent peaks we have scaled (the high correct classification results in *ideal authorship attribution* problems³⁹) we are essentially still at “base camp” in terms of understanding sociolectic stylistic variation in the GNT.⁴⁰ Moreover, it is important to clarify that each of these answers should be *linguistically* pursued, not just pursued with our favorite set of markers. That is, the entire set of questions above should be explored for each level of rank and at all levels of delicacy in the paradigmatic system, etc., and with enough *degrees of freedom* (e.g. a large enough set of variables and instances of the same) to proceed mathematically. This will generate likely thousands of experiments most of them

³⁷ By textual structure I mean the final linear form of the text. This is coextensive with the SFL notion of cotext.

³⁸ See especially the discussion of power-distance and high and low power-distance cultures in Brockner et al., “The Influence of Power Distance,” 301-02.

³⁹ See section 2.4.3.2 for a definition of ideal authorship attribution problems.

⁴⁰ Daeleman’s perspective is instructive here: “Computational stylometry, as in authorship attribution or profiling, has a large potential for applications in diverse areas... Yet, many of the basic research questions of this field are not studied systematically or even at all.” Daelemans, “Explanation in Computational Stylometry,” 451.

from “walking down” each system in the Koine *system network* to a level of delicacy that still leaves enough DOF for the analysis. (A calculation of the size of this set is presented in Chapter Four.) As one would surmise, such a task would require expanding the *automation boundary* of the analytic software used so that it can automatically perform these analytics via such things as *closed-loop software automation*, batch processing or the like.

2.3.3 Gaps in the Consistency of Findings

Lastly, given the dearth of multivariate studies in NT computational stylistics, it is perhaps little wonder that the state of computational stylistics in the GNT has so many open items—and that contradictions are rife within our discipline. To make that last point clear, with respect to the long-standing dialectic concerning Pauline pseudepigraphy, for instance, Morton asserted that Paul wrote only four epistles (the *Hauptbriefe*), Grayston and Herdan concluded ten (all epistles but the Pastorals),⁴¹ Gerard Ledger assigned six to Paul (including, intriguingly, 2 Thessalonians), Barr concluded six or seven, and Kenny concluded that twelve were “the work of a single, unusually versatile author.”⁴²

Seen abductively, then, given the research lacunae, contradictions, and sociolectic gaps just demonstrated, we see NT computational stylistics to be what it is—a relatively early stage discipline, yet one with many, many entry points for fruitful research.

⁴¹ To be fair, Grayston and Herdan didn’t exactly say Paul, but a first-century writer, who was presumably Paul. See the footnote immediately below.

⁴² Grayston and Herdan, “Authorship of the Pastorals,” 13–14; Morton, “The Authorship of Greek Prose,” 224–25; Barr, *Scalometry and the Pauline Epistles*, 125; Ledger, “An Exploration of Differences in the Pauline Epistles,” 95; Kenny, *A Stylometric Study of the New Testament*, 100.

2.4 Computational Stylistics: A Primer

2.4.1 Its Definition

The moniker “computational stylistics,” while no longer a neologism, is still fresh enough to have not yet gained a consensus definition in the literature. The most complete recent definition, in my view, is provided by Argamon and Koppel:

The computational analysis of the style of natural language texts, computational stylistics, seeks to develop automated methods to (1) effectively distinguish texts with one stylistic character from those of another, and (2) give a meaningful representation of the differences between textual styles.... Style, thus construed, includes interpersonal aspects of meaning such as affect, socio-linguistic categories such as genre and register, idiolectic aspects such as author identity and personality, and specifics of the individual speech act such as the medium of transmission and the purpose of the text.⁴³

In that same article, Argamon and Koppel later argue that their *communicative act model*⁴⁴ of computational stylistics is closely related to *systemic functional* models which more explicitly embrace *context of situation* and *context of culture* categories.⁴⁵ Hence, we would add only that their model seems less potentially productive in that it omits the sociocultural and *metafunctional categories* that a systemic functional model explicitly provides.⁴⁶ Despite this criticism, we submit the following two definitions for computational stylistics in order to build upon the definition offered by Argamon and Koppel. First, *ideally*, computational stylistics is the theory and praxis of quantitatively analyzing comprehensive or representative linguistic

⁴³ Argamon and Koppel, “Finding Meaning in Stylistic Variation,” 79–80.

⁴⁴ Argamon and Koppel, “Finding Meaning in Stylistic Variation,” 83.

⁴⁵ Argamon and Koppel, “Finding Meaning in Stylistic Variation,” 85.

⁴⁶ These remaining categories (or, in our preferred terminology, layers) omitted by these authors would include topic/topicality and subject matter, which is defined by SFL’s experiential metafunctional category. See especially Halliday and Matthiessen’s discussion of the experiential metafunction in Halliday and Matthiessen, *An Introduction to Functional Grammar*, 175-78. Argamon and Koppel in contrast, seem to view content as independent from other sources or causes of variation. From our perspective this is a distinction without a difference. Pragmatically and experimentally, subject matter categories are wholly defined by one or both systemic functional contexts, culture and situation, which in turn delimit the range of systemic choices made by the speaker/writer.

measures in a corpus using adequate and valid principles of experimental design for the purpose of answering a variety of linguistic, literary, social, contextual or authorial questions.⁴⁷ These questions explicitly embrace but are not limited to *sociocultural/contextually defined categories* such as genre/register, topic/topicality, subject matter; *idiolectic/authorially defined categories* such as author identity, personality, age, and gender; and *interpersonal categories* such as tenor, affect, persuasiveness, rhetorical categorization, and audience. Second, Argamon and Koppel elect to use the term computational stylistics to span the entire period of principled attempts to quantify style from the late nineteenth-century onwards. In my view “computational stylistics” should be limited to that period immediately following the period more commonly known as *stylometry*. Two reasons undergird this adjudication. First, large upheavals in theory and praxis have occurred in the discipline, and the term stylometry has, quite clearly, been sullied by its checkered past. Second, although many practitioners still call the discipline stylometry, many are beginning to transition to the term computational stylistics to denote that a sea-change has occurred in the discipline—one driven by both new methods (principally supervised and unsupervised multivariate analysis) and a new vigor that has flooded into the discipline from approximately the mid-1980s onward. Accordingly, this work views the transition from the period of stylometry to computational stylistics to be the mid-1980s, with the inflection point being the pragmatic and experimental transition from univocal to multiple causes of stylistic variation and the consequential use of more explicit methods of multivariate analysis, design adequacy, and statistical validation used to effect that transition.

⁴⁷ Notice we omit the requirement of automation that Argamon and Koppel include, while granting its necessity in praxis.

2.4.2 Its Recent Achievements

As just mentioned, the discipline experienced a sea-change in method, adequacy, and validity in the discipline beginning in the mid-1980s. Even more specifically, in the last twenty years, computational stylistics has, without question, made its greatest strides.⁴⁸ To sufficiently appreciate these achievements, we must provide more detail on computational stylistics than we presented in Section 1.2. Three trends are noteworthy.

First, no longer are language features (often called “markers” in the early history of stylometry) chosen *ad hoc*, by the mere caprice of the researcher. *Second*, practitioners have largely retired simple, serial univariate methods of textual analysis (chi-square and the like) in favor of multivariate methods.⁴⁹ These multivariate methods in turn, are divided into three categories: (a) discrete eigen-systems methods that decompose variation into components (a task, needless to say, very relevant to our goal of disentangling authorial from generic variation), (b) iterative, machine learning (ML) mathematics such as neural networks and (c) a rather heterogeneous grab bag of methods which include information theoretic approaches (IT) as well as information retrieval methods (IR). *Third*, these methodological advances have been coupled with various methods of validation. Each of these advances will be addressed in that order.

2.4.2.1 Recent Achievements: Methods

1. *More Representative or Outcome-Specific Methods of Picking Language Features*

(Markers) for Analysis. These approaches, often termed *feature set extraction methods*

(FSE)⁵⁰ or indexers, were largely developed in the last twenty years. Regardless of the

⁴⁸ Authorship attribution using quantitative measures is essentially coincident with the term stylometry. Perhaps no simpler definition of authorship attribution can be had than Grant’s: “Authorship analysis can be thought of as a classification problem in which one text, commonly referred to as the query text, is assigned to, or excluded from a group of texts.” Grant, “Quantifying Evidence in Forensic Authorship Analysis,” 3. Cf. to Juola’s slightly broader definition, Juola, “Future Trends in Authorship Attribution,” 120.

⁴⁹ More detail on multivariate methods will be provided in the primer addressing them (Section 3.9.3).

⁵⁰ More details on FSE will be provided in the primer on supervised multivariate analysis (Section 3.12.1).

approach, all these methods are similar in that they use one or more sets of mathematical criteria to select a subset of markers that meet some criteria (e.g. distributional, representativeness, randomness, etc.). That is, FSE enforces, in its most frequently implemented form, a kind of mathematical republicanism—selecting a few features to represent the entire population. Most often, that subset of features is selected so that it *discriminates best* for a given desired outcome (such as authorship or genre).⁵¹ This latter practice yields a down selection appropriate for the first task of computational stylistic—discrimination—but not for the second and third tasks, disentangling and causation respectively. FSE, in fact, almost guarantees a feature set that is not representative of the underlying population. In any event, by means of an analogy, FSE is an automated process that fuels the hungry furnaces of the “new stylistics”—computational stylistics.⁵²

2. *More Categories of Analytical Methods That Enable Greater Discrimination Between Authorship or “Other Layers.”* Four new categories of analytical methods have emerged from three separate points of the mathematical compass:⁵³

a. *Multivariate “decompositional mathematics”* (most typically *principal components, factor analysis, correspondence analysis, and discriminant analysis*).⁵⁴ These eigen-system-based methods have been used not only for

⁵¹ The former task is a *minimization* problem in mathematics, the latter a *maximization* (optimization) problem.

⁵² Our appropriation of the phrase “new stylistics” is intentional. It is an echo of the phrase first coined by Michaelson and Morton (Michaelson and Morton, “The New Stylometry,” *passim*). Their New Stylometry, however was not only not new—it used univariate methods just as before—it was retrograde, proposing, of all things, that style could be effectively transduced using a single lexical marker. Our intention is an attempt to rescue such a good eponym from such an undeserved fate.

⁵³ Frantzeskou et al. are correct in defining a taxonomy of four methodology categories: manual inspection, statistical analysis (including eigen-systems), machine learning and similarity measurement. We omit similarity measurement here because its methods (various forms of cluster analysis) present to the researcher neither the data reduction capabilities of unsupervised eigen-systems nor the utility of the automated feedback architecture of supervised ML methods. Accordingly, we will focus on Frantzeskou et al.’s middle two elements (Frantzeskou et al., “Source Code Authorship Analysis,” 90).

⁵⁴ See the excursus in Section 3.9.1 for more detail.

authorship disambiguation, but to tease apart idiolectic from sociolectic sources of variation (register/genre, subgenre, dating, gender, etc.).⁵⁵ These methods operate by mathematically extracting a small number of more interpretable components or “factors” from a larger number of original variables (in our case categories of a given linguistic measure). Accordingly, we will most often refer to these as *Extractive Multivariate Analysis (EMVA)* methods.

- b. *Machine learning (ML)*⁵⁶ includes iterative and convergent learning-based algorithms such as *neural networks, genetic algorithms, naïve Bayes, Support Vector Machines (SVM’s), Winnow, ensemble methods*, and the like.⁵⁷ These more computationally intensive and iterative techniques were made possible by dramatic increases in computational power over the past several decades. ML methods have most often been used for authorship disambiguation work and still remain a very active area of current research.⁵⁸

⁵⁵ See especially the initial studies by Biber and Burrows and the inaugural studies in the GNT by Neumann and Mealand in this regard. Burrows, “The Statistical Analysis of Narrative Style,” passim; Burrows, *Computation into Criticism*, passim; Biber, “Investigating Macroscopic Textual Variation,” passim; Biber, *Variation Across Speech and Writing*, passim; Neumann, *The Authenticity of the Pauline Epistles*, passim; Ledger, “An Exploration of Differences in the Pauline Epistles,” passim.

⁵⁶ For a terse summary of machine learning see Abbasi and Chen, “Applying Authorship Analysis to Extremist-Group Web Forum Messages,” 67–68. For a more comprehensive overview see Sebastiani, “Machine Learning in Automated Text Categorization,” passim.

⁵⁷ See specially the seminal studies of Abbasi and Chen, “Writeprints: A Stylometric Approach,” 7:4–7.19; Argamon et al., “Style Mining of Electronic Messages,” 476–78; Koppel et al., “Automatically Categorizing Written Texts,” 404–05; Koppel et al., “Authorship Attribution with Thousands of Candidate Authors,” 660; HaCohen-Kerner et al., “Identifying Historical Period and Ethnic Origin of Document,” 107–08; Stańczyk, “Application of DRSA-ANN Classifier,” 697–98; Tambouratzis, “Assessing the Effectiveness of Feature Groups,” 249–50; Tsimboukakis and Tambouratzis, “Neural Networks for Author Attribution,” 148–51; Tweedie et al., “Neural Network Applications in Stylometry,” 2–6.

⁵⁸ Special attention should be paid to the more recent work of Abbasi and Chen, “Visualizing Authorship for Identification,” passim; Argamon, “Scalability Issues in Authorship Attribution,” passim; Argamon et al., “Automatically Profiling the Author,” passim; Koppel et al., “Unsupervised Decomposition of a Document into Authorial Components,” passim; Koppel et al., “Authorship Attribution in the Wild,” passim; Stamatatos, “Authorship Attribution Based on Feature Set,” passim; Stamatatos, “Plagiarism Detection,” passim. Putniņš and Ebrahimpour, in particular, have executed initial ML work upon the GNT. Putniņš et al., “Advanced Text Authorship Detection Methods,” J1–J13; Ebrahimpour et al., “Automated Authorship Attribution,” 1–12.

- c. *Information retrieval (IR) and information theoretic (IT) methods.* These are principally deployed in authorship disambiguation studies. IR and IT, based on information compression and information content respectively, constitute rapidly evolving subfields within computational stylistics.⁵⁹
- d. *Latent Class Models.* In quantitative genre studies, an entire cottage industry has emerged that uses latent class models, typically forms of Bayesian models or topic models (such as the *latent Dirichlet allocation*, or *LDA*, model⁶⁰) to inductively discover boundaries between genres. Typically the resultant boundaries are circumscribed by inputting topics into the models across multiple genres.⁶¹ Because LDA has been used successfully to inductively discover genres in texts, its eventual use in the GNT is clearly indicated. In this present work, however, pride of place has been reserved for *a priori* notions of genre delivered to us by the more proximate, time-honored disciplines of literary criticism, genre criticism, and NT studies. The specific genre categories provided by the intersection of these disciplines will be further explicated in Section 5.4.

More thorough introductions to the EMVA, ML and IR methods can be inspected in Sections 3.9 and 3.15 respectively. Because latent class models will not be executed in this first cycle through our experimental design, a primer on its mathematics and praxis is beyond our current scope.

⁵⁹ See especially Marton et al., “On Compression-Based Text Classification,” passim; Koppel et al., “Authorship Attribution with Thousands of Candidate Authors,” passim; Zhao and Zobel, “Entropy-Based Authorship,” passim; Lambers and Veenman, “Forensic Authorship Attribution Using Compression Distances,” passim; Luyckx, *Scalability Issues*; Mastroiolo et al., “Legal Documents Categorization by Compression,” passim; Oliveira et al., “Comparing Compression Models,” passim.

⁶⁰ This modeling, in our view, is especially appropriate for application in the Koine. See the seminal mathematical instantiation of LDA in Blei et al., “Latent Dirichlet Allocation,” 993–1015.

⁶¹ Researchers developed topic and text classification modeling as an outgrowth of *web-informatics*, the need for researchers, companies and to individuals to classify web pages, blogs, and other web content by category.

3. *More Focused Methods of Validation*: Typically, advanced validation procedures operate in concert with supervised methods and FSE. In its simplest form, validation is simply a process that ensures that the classifier produces results above chance alone. It does so by ensuring that the classifier can reproduce *known* texts correctly before it can be entrusted with the true riches—if a Lukan metaphor can be permitted—of classifying *unknown* texts. Four conceptual steps are involved: (1) the ML/IR classifier makes an initial, often random change as to how it weights, balances or combines the features (linguistic measures) that it uses to develop its classification. (2) The validation procedure holds back some of the known members (e.g. texts written by a known author) from the set that the ML/IR classifier is being trained upon. (3) The fully trained ML/IR classifier is now run upon the hold-back sample to develop some measure of goodness of fit (e.g. such as percent correctly classified). (4) The prior action taken to train the model is evaluated. If the outcome is improved from the last iteration, that change is retained, if not, it is omitted. The process is repeated until no more improvement can be made to the ML/IR classifier. Many specific kinds of validation exist such as cross-validation, k-fold cross-validation, jackknife validation, and the like.⁶²

⁶² Validation is now *de rigueur* for supervised multivariate classification tasks. While all methods of validation share the core similarity defined by the conceptual steps listed above, inspection of the literature reveals three overlapping phases in the use of validation in computational stylistics. In the first phase, simple *crossvalidation* methods were employed in the early 1990s (e.g. Matthews and Merriam, “Neural Computation in Stylometry I,” 205; Holmes and Forsyth, “The Federalist Revisited,” 117). Second, animated no doubt by the seminal and principled comparison article by Kohavi (Kohavi, “A Study of Cross-Validation,” 1137–43), increasingly deep *k-fold validation* began to emerge by the mid- to late-1990s (e.g. Tweedie et al., “Neural Network Applications in Stylometry,” 5–6; Dietterich, “An Experimental Comparison of Three Methods for Constructing Ensembles,” 140; Koppel et al., “Automatically Categorizing Written Texts,” 405). Third, more sophisticated varieties of validation such as *jackknife validation* (leaving out observations one or more at a time) as well as more exhaustive combination-based methods emerged by the late 1990s. For jackknife validation see the early work of Kjell (Kjell, “Authorship Determination Using Letter Pair Frequency,” 123; Kjell, “Authorship Attribution of Text Samples,” 1660) and the latter work of Mosteller (Mosteller and Wallace, *Inference and Disputed Authorship*). For the earliest instance of validation, albeit with the an older, non-machine learning classifier (multiple discriminant analysis), see Mealand, “Measuring Genre Differences in Mark,” 236.

2.4.2.2 Recent Achievements: Praxis

In the last three decades two primary modes of multivariate praxis have been used to address computational stylistic problems: *unsupervised and supervised multivariate methods*. In one sense, these two modes can *seem* similar because they both can be used to achieve the same end: classification of an unknown text into a given category, whether that category be authorial, generic or otherwise. Yet, there the similarity ends, for these modes go about their business in very different ways, and (with one exception) the mathematics they employ are quite distinct. Let's return to our "river metaphor" in Chapter One to further define, compare and contrast them.

The first river to scour the stylistic landscape was *unsupervised multivariate analysis* made possible by *eigen-system mathematics* (i.e. EMVA methods). In stylistics problems EMVA starts with the tangle of language measure(s)⁶³ always present in natural language texts and draws out from the chaos multiple coherent sets of components—that is, categories within those language measure(s) that are self-similar to one another yet uncorrelated to other components.⁶⁴ These components can then be assayed for their relationship (usually by mathematical association but also by correlation or by latent factor analysis) to underlying *causes* of that variation which researchers have identified as authorship, genre, date, gender etc. This river best explains the *structure* in corpora without prejudice because it leaves the structure intact, making the epiphenomena (its layers or causes) more visible by washing away the obscuring mud. In its most mature instantiation (latent structure analysis), it lays bare which smaller roots (observed

⁶³ Or better yet, within a given *language measure*, where language measure means a coherent linguistic complex, whether syntagmatically, paradigmatically or extralexically-defined (measures from another language strata.)

⁶⁴ That is, eigen-systems methods extract from the total variance successive dimensions that are uncorrelated with prior dimensions. Literally, after the first and most important dimension is identified, that variance is *extracted* from the total variance and the residual "what's left over" is resubmitted to the same process iteratively until no remaining variation is left.

language categories) feed into more primary roots (unobserved components). Three of the four primary EMVA methods—principal components analysis, factor analysis, and correspondence analysis—are the frontline tools used in unsupervised multivariate analysis.⁶⁵

The second river to scour the stylistic landscape was *supervised multivariate analysis*, typically instantiated through the use of *machine learning (ML) and information retrieval (IR) mathematics*. This is the partner in the methodological duo that has garnered the more recent and sensational headlines.⁶⁶ In contrast to unsupervised approaches, supervised approaches do not focus on the structure of the linguistic landscape—this it obliterates. Rather, supervised approaches are interested in one thing—classification. To classify unknown texts they extract the linguistic gold buried within that landscape. Supervised approaches can extract that gold efficiently for three reasons. First, they are armed with something unsupervised methods lack: *prior information*. In the marquee example of authorship attribution, that prior information is known texts—texts the researcher is essentially certain to have been authored by a given individual. Second, as the classification operation begins, it starts with all the linguistic raw material just as (ideally) unsupervised multivariate analysis does. There, however, the similarity ends. Rather than keeping all the raw linguistic material and analyzing it to discover its covariance structure (its relative association with all other categories), supervised mathematics

⁶⁵ The meta-analysis executed for this work identified 152 computational stylistics studies that used eigen-systems mathematics. 21 of those were EMVA studies executed upon the biblical texts. Early or important studies among these include Radday et al., “Genesis, Wellhausen and the Computer,” 470–74; Radday and Shore, *Genesis: An Authorship Study*, 93–121; Neumann, *The Authenticity of the Pauline Epistles*, 153–66; Ledger, “An Exploration of Differences in the Pauline Epistles,” passim; Mealand, “The Extent of the Pauline Corpus,” passim; Linmans, “Correspondence Analysis of the Synoptic Gospels,” 4–11; Mealand, “Style, Genre, and Authorship in Acts,” 484–99; Putniņš et al., “Advanced Text Authorship Detection Methods,” J6–J12; Mealand, “Is there Stylometric Evidence for Q?,” 487–99; Erwin and Oakes, “Correspondence Analysis of the New Testament,” passim; Mealand, “Hellenistic Greek and the New Testament,” 328–42; Ebrahimpour et al., “Automated Authorship Attribution,” 10–11.

⁶⁶ Our meta-analysis is still relatively incomplete for machine learning methods. Our work to date has nonetheless identified seventy-one computational stylistics studies using machine learning/IR. If we include extra-textual studies (e.g. on web pages, XML, source code etc.) the number increases to 116 studies. Intriguingly only one of those studies is upon the biblical texts (Ebrahimpour et al., “Automated Authorship Attribution,” 6–11).

tests each component (e.g. each word at the rank of lexis or each clause at the rank of the clause) either alone or in combination with other categories. If that category does not improve the classification, the foreman of the operation (the classifier algorithm) peremptorily throws it on the scrap pile (or de-weights its contribution). Third, supervised methods, almost without exception, are automated in a closed feedback loop with FSE and measures of validation. Thus equipped, it is not surprising that supervised methods excel at what authorship attribution researchers, at least, consider to be the central task of computational stylistics: classification. In classical textual studies (in contrast to web-informatics) supervised methods disproportionately use the ML/IR classification engines described above.⁶⁷

2.4.2.3 Recent Achievements: Results

Most importantly, these new methodologies, their modes, and novel applications have yielded, in no uncertain terms, a new era of improved results. Four such results are noteworthy:

1. *Results from Feature Set Extraction* (FSE) methods. By more representatively estimating the underlying distribution of linguistic measures (features) in texts, FSE has yielded three benefits. First, when deployed with EMVA methods, the use of fewer and more representative features reduces the likelihood of phenomena that confound eigen-systems-based multivariate analysis such as *multicollinearity*.⁶⁸ Second, when used with EMVA methods, a reduced feature set produces a solution of lower dimensionality when mapped or graphed. This allows the characteristic bi-plots from EMVA to account for more of the total variance across all dimensions of the eigen-system solution (~1990 to

⁶⁷ One form of EMVA mathematics, discriminant analysis (DA), is supplied with prior information. Hence, especially in the early decade of the rise of supervised approaches, DA became the primary method for performing supervised multivariate analysis upon texts.

⁶⁸ Multicollinearity occurs when identical or nearly identical features prevent eigen-systems mathematics from proceeding at all, or from producing stable results.

the present). Third, when executed (as is most frequently the case) in concert with supervised ML or IR methods, FSE operates in a closed loop feedback system to rapidly improve results until a maximum achievable fit is achieved. See point three below.

2. *Results from EMVA Methods*: EMVA methods have successfully and parsimoniously separated language into components associated with long-standing hypothesized causes such as genre, dating, gender, style-shift, and the like. This has lent empirical support for the kinds of things literary and linguistic theorists such as Hirsch, Bakhtin, Hasan and Halliday have been telling us were characteristic of texts all along⁶⁹ (from the mid-1980s to the present).
3. *Results from Combining FSE, Supervised Methods and Validation*. Since the 1980s in particular, this integrated triad of technologies has been used to solve whole classes of problems that had previously eluded researchers or yielded suboptimal outcomes. The success of these technologies, moreover, has caused their application to expand far beyond authorship attribution alone to embrace genre,⁷⁰ gender,⁷¹ dating, and stylochroometry problems,⁷² as well as to yield other literary and web-informatics

⁶⁹ For contemporary literary theory regarding genre, see Bakhtin, *Speech Genres and Other Late Essays*, 60–67; Hirsch, *Validity in Interpretation*, 68–126; Aune, *The New Testament in its Literary Environment*, 17–157. For contemporary linguistic theory on text-types/genre/register see Hasan, “Text in the Systemic-Functional Model,” 229–31; de Beaugrande, “Register...in Search of a Theory,” 7–13; Halliday, *On Grammar*, 229–30; Halliday, *On Language and Linguistics*, 195.

⁷⁰ See Table 2.2 in Section 2.4.3.1. We count fifty-seven such multivariate genre studies with the earlier work establishing the success of the separating texts by genre. See especially Carroll, “Vectors of Prose Style,” 283–92; Brainerd, “On the Distinction between a Novel and a Romance,” 259–70; Faigley and Meyer, “Rhetorical Theory and Readers’ Classifications of Text Types,” 305–26; Burrows, “The Statistical Analysis of Narrative Style,” 61–70; Burrows, *Computation into Criticism*, 153–75; Biber, *Variation Across Speech and Writing*, 61–97; Burrows and Craig, “Lyrical Drama,” 63–86.

⁷¹ The meta-analysis performed for this study currently numbers seventeen multivariate gender classification studies. See especially Koppel et al., “Automatically Categorizing Written Texts,” 402–07; Argamon et al., “Gender, Genre, and Writing Style,” 325–42; Kucukyilmaz et al., “Chat Mining for Gender Prediction,” 276–328; Argamon et al., “Stylistic Text Classification,” 806–16; Cheng et al., “Author Gender Identification from Text,” 80–86; Suero Montero et al., “Investigating the Role of Emotion-Based Features,” 101–10.

⁷² See especially Temple, “A Multivariate Synthesis,” 67–72; Forsyth et al., “Investigating the Authenticity of the Consolatio,” 378–94; Can and Patton, “Change of Writing Style with Time,” 66-76; HaCohen-Kerner et al.,

outcomes.⁷³ Most importantly, the synergistic nature of this combination has produced a remarkable leap forward in classification accuracy. Under the proper design conditions (and with a compliant data set) correct classification percentages or lift results have frequently exceeded 90 percent of the theoretical maximum⁷⁴ (mid-1990s to the present).

4. *New Categories of Automated Software* have been developed (such as WEKA, JGAAP and the like).⁷⁵ These tools contain dozens or even hundreds of discrete mathematical algorithms that can be executed to discover the single classifier, FSE and/or validation mode that maximizes correct classification for a given problem (2000 to the present).

“Identifying Historical Period and Ethnic Origin of Document,” 106–11; HaCohen-Kerner et al., “Stylistic Feature Sets as Classifiers,” 852–59.

⁷³ Important examples of the inaugural work in web-informatics include the following: Chan and Chong, “Unsupervised Clustering for Nontextual Web Document Classification,” 377–96; Park and Zhang, “Co-trained Support Vector Machines for Large Scale Unstructured Document Classification ” 421–39; Abbasi and Chen, “Applying Authorship Analysis to Extremist-Group Web Forum Messages,” 67–75; Guo et al., *A Fast Document Classification Algorithm*, 186–89; Abbasi et al., “Stylometric Identification in Electronic Markets,” 49–78; Abbasi et al., “Sentiment Analysis in Multiple Languages,” 1–34; Berry et al., “Document Classification Using Nonnegative Matrix Factorization,” 2782–85; Kanaris and Stamatatos, “Learning to Recognize Webpage Genres,” 499–512 and Mason et al., “Classifying Web Pages by Genre,” 458–65

⁷⁴ See, for instance, the studies by Abbasi and Chen, “Applying Authorship Analysis to Extremist-Group Web Forum Messages,” 72; Stańczyk, “On Performance of DRSA-ANN Classifier,” 178; and Marsden et al., “Language Individuation and Marker Words,” 7. It must be understood, however, that generally classification cannot be fairly compared between studies because factors such as set similarity, the number of discrete sets, and the goal of prediction (e.g. authorship or genre) usually differ significantly. Better cross-study comparators do exist such as lift (percent of the difference between chance and the theoretical maximum) but no comparator is wholly without bias. Our point, however, still stands. The era of supervised mathematics has dramatically improved our ability to correctly classify documents. For concerns about the reporting of some of these studies, see Luyckx and Daelemans, “Authorship Attribution and Verification with Many Authors and Limited Data,” 519.

⁷⁵ Currently eighty-five reviews or studies in our meta-analysis mention or use WEKA (Waikato Environment for Knowledge Analysis) software, a freely available machine learning software package originally introduced by Witten and Frank; Witten and Frank, *Data Mining: Practical Machine Learning Tools*, 265–320. For the earliest WEKA-enabled work see especially Finn et al., “Genre Classification and Domain Transfer,” 361; Berger and Merkl, “A Comparison of Text-Categorization Methods,” 1000; Luyckx and Daelemans, “Shallow Text Analysis and Machine Learning,” 151; Zhao and Zobel, “Effective and Scalable Authorship Attribution,” 182. JGAAP is a similar learning based software package but is designed as an end-to-end software framework for solving authorship attribution tasks per Juola et al., “A Prototype for Authorship Attribution Studies,” 173–74. JGAAP appears to be less widespread in use than WEKA with nineteen articles or reviews in our bibliographical database. Some notable work done with JGAPP includes the following; Juola and Vescovi, “Empirical Evaluation of Authorship Obfuscation Using JGAAP,” 14–18; Juola and Vescovi, “Analyzing Stylometric Approaches to Author Obfuscation,” 115–25; Juola, “Detecting Stylistic Deception,” 91–96; Brennan et al., “Adversarial Stylometry,” *passim*; Stuart et al., “On Identifying Authors with Style,” 3048–53 and Stolerman et al., “Breaking the Closed-World Assumption,” 185–205

How might we summarize the effect that these changes in praxis produced in the discipline of computational stylistics? As far back as 2003 (which is a long time in this discipline), Holmes signaled the early sea-change these methods were beginning to bring to classic stylometric work:

Following the seminal work of J.F. Burrows (1992), multivariate statistical analyses involving large sets of (50-100) of non-contextual features have met with astonishing success, recent examples being the study on the *Provenance of Doctrine Christiana*, attributed to John Milton (Tweedie et. al. 1998) and the attribution of the 1583 *Consolatio*, shown to be not a lost work of Cicero but a sixteenth century forgery (Forsyth et. al. 1999).⁷⁶

Given, as we have seen, that well-constrained ideal authorship attribution (IAA)⁷⁷ studies can typically achieve correct classification accuracies above 90 percent, one would think that our job here is done. Haven't we been translated, essentially, to a sort of stylistic nirvana? What more is there left to do? It is here where we must be clear in our critique. Clearly, high-performance point solutions to some vertical problems in stylometry have been developed.⁷⁸ These, however, relate almost entirely to the first class of problems in computational stylistics, *classification*. We have not made anywhere near that kind of progress in the other two classes of

⁷⁶ Holmes et al., "Stylometry and the American Civil War," 406. One of the central figures behind this renaissance, Stamatatos, in reviewing the field concludes that in the last decade "this scientific field has been developed substantially, taking advantage of research advances in areas such as machine learning, information retrieval, and natural language processing." Stamatatos, "A Survey of Modern Authorship Attribution Methods," 538.

⁷⁷ Ideal Authorship Attribution (IAA) is a concept borrowed from Hoover (Hoover and Hess, "Authorship Attribution: the Mysterious Maria Ward," 467-68). IAA studies, in our view, are those in which four conditions are simultaneously met: (1) The corpus contains a large set of known texts from distinct authors. All else being equal, this enables the training set to yield excellent separation between known authors (small intra-authorial variation and large inter-authorial variation). (2) A feature set has a large set of types and tokens from which to draw, preferably across the rank scale of the language. (3) The availability of multiple supervised mathematical methods executed upon the feature set(s). (4) The ability to combine model sets together via arbitration (voting). This improves overall classification by identifying the conditions in which one model outperforms another when they disagree. Per Koppel et al., most authorship attribution problems, however, do not provide the researcher with a set of conditions this ideal. Koppel et al., "Computational Methods in Authorship Attribution," 9.

⁷⁸ This is true provided that unexplored covariance between causes has not confounded the identifications. That is, even though the newer supervised point solutions separate the sets of known texts efficiently and yield higher correct classifications, but they are just as liable to commit casual misassignment if there is strong covariance between one cause (e.g. authorship) and another (e.g. genre). See The Experimental Problems in this section.

problems in computational stylistics: *structure* and *causation*. The lack of relative progress regarding these last two problems, in my view, are still profound enough so as to prevent the deepest potentials of the discipline from being realized. And there, borrowing Dreher's apt phrase, hangs a tale.⁷⁹ Accordingly, two final introductory issues must be considered: the shifts and trends in computational stylistics and the still unsolved problems in the discipline.

2.4.3 Its Problems

2.4.3.1 The Imbalances in Published Studies

What, then, are the still unsolved problems in the discipline? One way to get our arms around the problem is to gain a sort of "aerial overview" of the discipline.⁸⁰ This was achieved by crosstabulating the current body of 237 published multivariate computational stylistics studies in our database by method and objective. This yielded the results in Table 2.2.⁸¹ Each cell of the table reflects a study (or part of one) in which a given method (in the rows) was used to pursue a given objective (in the columns). Notice that if a study executes, for instance, three methods or produces results across multiple objectives (e.g. operates upon multiple sources of variation) it will be counted for each outcome or method.⁸²

⁷⁹ Dreher, "The Computer-Linguistic Detective of Authorship," 883.

⁸⁰ We borrow this notion from Kenny. See especially Kenny, *A Stylometric Study of the New Testament*, 116.

⁸¹ Table 2.2 subsumes literature up through November 2014. While no metaanalysis can claim to be perfectly exhaustive, this effort appears to contain roughly the same depth of references as the bibliography of Rudman (1,500 entries). See Rudman, "The State of Non-Traditional Authorship Attribution Studies-2012," 265. (The entries for this metaanalysis stand at 1,552 but that includes corpus linguistic and relevant mathematical studies.)

⁸² Also note that ML-based methods executed on web content have been largely excluded, thus limiting the summary totals in Table 2.2 to those studies executed upon traditional texts. Had I included recent web-informatics-studies related to text classification it is likely that four times as many studies would have been included in those cells.

**Multivariate Computational Stylistic Studies Overall⁸³
By Objective and Method**

			Objective / Style Subcomponent Analyzed					Totals
			Idiolectic		Sociolectic			
			Authorship	Genre/ Register	Date of Composition	Gender	Others	
Method of Analysis	EMVA-Based	PCA	56	17	4	2	3	68
		FA	5	9	1	5	2	16
		MDA	65	1	1	3	22	83
		MDS	0	1	0	0	3	4
		CA & MCA	5	3	1	0	2	9
	Distance-Based	Cluster Analysis	41	6	1	0	15	61
	ML-Based	Neural Net	22	2	0	0	1	25
		SVM	22	13	4	6	6	27
		IR	5	2	0	0	0	7
		Naïve Bayes	8	0	1	0	1	10
Totals ⁸⁴			169	57	11	17	29	237

Table 2.2

At the highest level, Table 2.2 yields four distinct findings and five trends. First, in terms of findings, the meta-data in Table 2.2 necessitates, at the very least, that we redefine the sources and nature of *stylistic variation* in texts. Specifically, whereas the preponderance of prior univariate studies presumed that stylistic variation *was only idiolectic in nature* (related to the individual) with authorship as its primary or sole component, the multivariate studies in Table 2.2 collectively demonstrate that (1) sociolectic (socially-related) stylistic variation *exists*, (2) its components include, at a minimum, genre/register, date of composition, and gender and (3) other components *may* exist depending upon the nature of the texts included in the corpus under study.⁸⁵ In sum, these 237 studies allow us to propose a working definition of stylistic variation as follows:

⁸³ Although the effort inaugurated here is a multi-year project, I have summarized in Table 2.2 only the high-level findings related to the specific concern of this study, the relationship of sociolectic to authorial variation in texts.

⁸⁴ Note that because a given study may produce data for multiple outcomes the row totals may be less than the number of studies in the individual rows that constitute that column.

⁸⁵ It is to be noted that the categories and subcategories of sociocultural variation differ somewhat between literary and linguistic theorists, and differ again between those who have studied the data inductively “up from corpora.” Complicating this is that no consensus enumeration of the complete set of sociologically-related (sociolectic) or individually-related (idiolectic) sources of variation for the GNT currently exists among either NT scholars or among computational stylists. It is for this reason that we have proposed the working definitions in this section. For a definition of sociolect from the perspective of modern computational stylistics see Juola, “Future Trends in

Working Definitions of Stylistic Variation

1. The total stylistic variation in a heterogeneous corpus of texts (such as texts from a mix of genres) is minimally comprised of two sources; a sociolectic (sociologically-related) source and an idiolectic (individually-related) source.
2. The *sociolectic component*, includes, at a minimum, the subcomponents of genre/register, date of composition, target audience in terms of culture, target audience in terms of one person or many (number), and social status of the author or recipient.
3. The *idiolectic component* includes, at a minimum, the subcomponents of author identity (authorship), but may also include author personality type, age of author, and gender of author.⁸⁶
4. Given the above definitions, *genre/register (our concern in this study) is therefore a subcomponent of sociolect, and authorship (our other primary concern in this study) is a subcomponent of idiolect.*

The Importance of This Working Definition of Sociolect and Idiolect

The importance of this working definition of *idiolect* and *sociolect* is threefold. *First*, it deals responsibly with the computational stylistic data of today without going beyond it. That is, the extractive (nonsupervised) multivariate work summarized above tells us, in one sense what we already know. There are *individually-related* causes of stylistic variation (what we term here idiolect) and there are *sociologically related* causes (what we term here sociolect) and within

Authorship Attribution,” 4. Cf. with Harrison’s view of the same; Harrison, *The Problem of the Pastoral Epistles*, 48–66.

⁸⁶ The construct of dialect which clearly contributes to stylistic variation may arguably fit into either sociolectic or idiolectic categories. In our view it is defined by the context of culture and hence is closer to idiolect. For the time being we will leave it out of our consideration here.

each there are an indeterminate number of components. We also know that authorship is a component of the former and genre is a component of the latter. Beyond this we cannot say all that much more experimentally—*yet*. This leads to our second point. *Second*, supervised multivariate analysis, though is garnering all the headlines is *not* designed to get us to the fundamental *structural* understandings of layers or components of stylistic variation in corpora. An analogy is helpful here. The discipline of geology and mineralogy provide just such an analogy. State-of-the-art geology (Some of my colleagues are geo-physicists) is a relatively advanced science. The modern geologist can look at almost any geographical structure on the *surface* of the earth and (s)he can relatively accurately describe what is *underneath* the surface as well as the composition (components) of those subsurface layers. How is this possible? The shapes of visible terrain *coupled with the advanced tools of the geophysicist* (“*thumper trucks*,” *subsurface radar, core samples, etc.*) reveal the hidden structure underneath the surface undulations. The tools of computational stylistics that are akin to the tools of geophysicists are the various flavors of unsupervised multivariate analysis just discussed (factor analysis, PCA, and various forms of structural equation modeling). These unsupervised mathematical tools thump the linguistic ground to tell computational linguists about the hidden layers underneath the visible undulating surface variation in texts. The problem? This kind of work is, comparatively, at a far earlier stage than the work of geophysicists. We need more “thumper truck” studies (more unsupervised multivariate analysis run upon mixed and unmixed corpora) for scholars to begin to know what layers of variation *consistently* underlie the surface variation in texts. *Third*, and last, there is the analogy of the *mineralogist*. The central concern of the mineralogist is to dig out of the earth its motherlode, and to use those precious minerals for his or her own (ostensively beneficial) purposes. If a generalization can be countenanced, the mineralogist doesn’t really

care about the layers they need to get to (or through) in order to get at the vein that holds their treasure. Mineralogists, in our analogy, are like computational stylists who use *supervised multivariate analysis* (recall Section 2.4.2.1) in their research. They are not concerned with underlying structure of variation. They are looking for classificational “gold”. They start with all the variation but throw out any paradigmatic, syntagmatic or semantic data that does help them predict the “gold” they seek; that is, the outcome (often authorship attribution) they are trying to predict.

Trends

Several trends can be drawn from this “aerial overview” when combined with the rate of change in those publications over time:

- Historically about fifty percent more idiolectic studies have been performed compared to sociolectic studies (169 versus 114).
- The publication rate of sociolectic studies over time (especially document classification) is increasing while idiolectic studies are staying approximately constant.⁸⁷
- The publication rate of EMVA studies (especially unsupervised studies) is slowing while supervised ML studies are rapidly increasing.
- The publication rate of traditional textual-based studies is relatively constant while the web-content-based studies appear to be rapidly increasing (blogs, emails, social media, Twitter etc.).⁸⁸
- The largest rate of increase in sociolectic studies are in the subcategory of web document classification (topicality, sentiment analysis, extremist group membership, etc.).

These recent shifts and trends display four sets of imbalances in the published literature.

(1) Historically, more idiolectic studies than sociolectic studies have been performed. (2)

Unsupervised studies, *the very studies that take into account all the data of the text* and yield its

⁸⁷ Most of the increase in the publication rate in sociolectic studies is due to web-document text classification studies.

⁸⁸ Using very restrictive criteria (limiting our search to studies whose titles contain one or more of these words: web, blog, extremist, e-mail, hate, forum, twitter, Facebook or social media) only thirteen computational stylistic studies were published in the seven years between 2001-2007 and forty in the seven years between 2008-2014.

fundamental components and structure, are in *decline* relative to supervised studies. (3) There is a relative absence of Latent Structural studies in textual (as opposed to web-content-based) studies. These are the kinds of studies that should be increasingly pursued if we seek to understand causation in texts. (4) There is a relative absence of nonparametric multivariate methods (e.g. log-linear models) in recent research. Little can explain this neglect given that core data (whether it be textual or metatextual) is fundamentally frequency-based and understanding the interactions between genre and authorship are a clear *desideratum*.

2.4.3.2 The Experimental Problems

Two primary experimental problems concern us: inadequately designed experiments, and the disconnect between supervised and unsupervised studies. We will treat each in turn.

2.4.3.2.1 Inadequately Designed Experiments

Section 2.1 demonstrated that reviewers of stylometry such as Holmes, Dixon, Rudman and Juola have heavily critiqued the propensity of the earlier studies to adopt particularistic *ad hoc* measures analyzed with *ad hoc* mathematics, only to arrive at a set of conclusions that are verifiably contradicted by other combinations of measures and methods. In the parlance of the experimentalist, such contradictions can only occur when the conclusion(s) of at least one of the experiments being compared constitutes a *false positive* or a *false negative*. Thus, Type I and Type II errors still dog us even in our current, vastly more technologically-enabled era of supervised and unsupervised studies. This underscores the need for careful and rigorous experimental designs. This issue will be further addressed in section 2.6.2.1.

2.4.3.2.2 The Unsupervised/Supervised Antinomy

Second, once we dig into it, there is an implicit antinomy in the collective computational stylistics findings in Table 2.2, although it is not apparent at first glance. Recall that supervised

methods operate by developing an optimal subset of features that best separate some desired outcome (authorship, genre, date, etc.). *There is however a cost to developing such a finely tuned subset of “markers.”* Because supervised methods are focused on one thing (say, discriminating between Paul and not Paul) they (using our river metaphor again) scour away every linguistic measure not best suited for the task.⁸⁹ In so doing, supervisory methods throw out the vast preponderance of linguistic *variation*—and hence *covariation*—necessary to disentangle the various causal layers in texts. In short, supervised methods throw out the structural linguistic baby with their supervisory bathwater.

So whence the antinomy? Unsupervised methods tend to do more poorly at classification for a given causal outcome (whether it be idiolectic, sociolectic, dialectic, etc.) but they keep all the linguistic raw material fed to them and attempt to understand the position and connections each linguistic piece has to the rest of the causal landscape. Supervised methods, in contrast, do better at classification but care nothing for the landscape with its deeply nested linguistic ecosystems. Rather they strip mine it to extract the gold (the optimal feature set) demanded of them by their overlords, the prior information fed to them. This prior information allows the researcher to say in effect to the supervised method, “Here is the outcome. Now feel free to weight, recombine or throw out anything in order to achieve this outcome.” Such unconstrained freedom, *ceteris paribus*, is far more likely to produce a classification model that is less generalizable (overfit)—especially when (a) the method is not thoroughly validated, (b) the *feature set* selected is poorly representative of the entire *linguistic potential* encoded within the

⁸⁹ The newer form of supervised approaches do not select their linguistic measures (features) *ad hoc*, but automatically. They do so by iteratively arriving at a final set of features that maximizes an “objective function”—some desired outcome such as overall correct classification.

different author sets, or (c) when, again, there is strong covariance between causes.⁹⁰ In the latter case, our vaunted and advanced supervised methods may be just as liable to Type I casual error (e.g. that is, assigning the cause of a trained set to cause X, when in fact it is due to cause Y⁹¹) as in prior eras populated by less powerful methods. The implication of this antinomy, at least in terms of the very relevant issue of the authorship of the GNT (or any mixed genre corpus) is quite clear. Prediction of any desired outcome (authorship or genre etc.) in computational stylistics *must be an exercise in both supervised and unsupervised classification*. Unsupervised classification or other extractive methods must be used to “disentangle” the layers of variation in texts so that a supervised separation “stays honest.” That is, when supervised classifications are executed as they typically are, (*upon the total variation in a corpus rather than upon its components*) supervised approaches can misidentify variation as being due to cause X when in fact it is due to cause Y. This variety of Type I causal error is especially prevalent when researchers have for any reason discounted cause Y (or perhaps not even thought of it) and *especially* if Y turns out to constitute more of the total variation being measured than X. The central question of this dissertation, in fact, is to test whether certain studies in the history of NT scholarship have unintentionally confounded X (authorship) with Y (genre). Now that the first primer, computational stylistics, has been discharged, it is time to move onto the second planned primer, the critical and central role linguistics must play in any comprehensive and experimentally disciplined approach to computational stylistics. This we do next.

⁹⁰ This antinomy, and the desire to combine the benefits of both supervised and unsupervised methods, is reflected in Layton’s research which attempts to “address the demand for an automated and unsupervised method for Authorship Analysis.” See Layton et al., “Automated Unsupervised Authorship Analysis,” 97.

⁹¹ Where Y can be one or more causes.

2.5 Linguistics: A Primer

2.5.1 The Motivation for Studying the NT Linguistically

After the turn of the millennium, researchers such as Argamon, Whitelaw, and Koppel began aggressive research programs to mine their corpora using formal or functional measures from various schools of linguistics.⁹² This more patently linguistic perspective exposed that while over 150 unsupervised multivariate analysis studies have been executed since the mid-1980s, *virtually all of these have been executed using only measures drawn from traditional grammar*. At least from the standpoint of measures based on the various schools of modern linguistics, then, we really don't know that much *linguistically* about any texts, including the GNT. We have already enumerated the sociolectic gaps regarding the GNT in section 2.3.2. To these we add other fundamental questions informed by an explicitly linguistic perspective such as:

- How do the texts of the GNT cluster by the various linguistic strata, ranks or systems of the Koine? Is it by authorship first? Genre? Something else?
- How does that clustering differ when the most discriminatory measures are used for each 'layer' (e.g. authorship, genre etc.)? How does it differ when supervised or unsupervised methods are used?
- Is there a pattern to how the various linguistic measures (strata, ranks, or selection expressions within a system) map to putative underlying idiolectic or sociolectic causes?

⁹² See especially Argamon and Dodick, "Corpus-Based Study of Scientific Methodology," 222–28; Whitelaw et al., "Identifying Interpersonal Distance," 204–12; Argamon et al., "Stylistic Text Classification," 803–17; Argamon et al., "Language Use Reflects Scientific Methodology," 215–33; Koppel et al., "Computational Methods in Authorship Attribution," 15–22.

- Are low-frequency or high-frequency linguistic measures more associated with a given cause or layer?
- To what extent does diglossic variation affect these questions? Or dialectal variation? What about variation due to the disparate facility (or lack thereof) in the Koine by the various authors/editors/redactors of the text?

Other gaps in our knowledge can be listed here, but these alone demonstrate that we have, at best, a piecemeal view of stylistic variation in the GNT, rather than a fundamentally linguistic one, and a more patently linguistic approach is indicated.⁹³ Given these gaps, how might a linguistic perspective aid our central objective of disentangling authorship from genre? To answer that question we will focus on only two central topics: (1) clarifying the relationship between traditional grammar and linguistics and (2) laying out a rationale for using Systemic Functional Linguistics in particular.

2.5.2 The Slow Linguistic Turns in NT Studies and Computational Stylistics

Virtually all NT scholars would grant that the still reigning paradigm of Koine Greek pedagogy and analysis is traditional grammar. Yet, unlike a generation ago, very few contemporary NT scholars would argue that that hegemony still stands unchallenged. A recent review of the literature by Porter and Pitts enumerates almost 200 peer-reviewed articles, monographs or books arguing that modern linguistics provides a more adequate model for understanding the language of the GNT than traditional grammar and, consequently, linguistics is critical to the synchronic analysis, exegesis, and interpretation of the Greek NT.⁹⁴

⁹³ Nini would go further by claiming that linguistics, specifically SFL, is capable of bridging the divide between the cognitive and stylistic debate in forensic authorship analysis (Nini and Grant, "Bridging the Gap between Stylistic and Cognitive Approaches to Authorship," 174–99).

⁹⁴ The beginning of this thinking can be traced more than anything else to Barr's ground-clearing work in 1961. In his *Semantics of Biblical Language*, Barr demolished the naïve constructivist approach to semantics which viewed interpretation as merely a serial exercise in chaining together word studies from Kittel. See Barr, *The Semantics of Biblical Language*. Instead Barr argued for a synchronically oriented, linguistically informed approach to language.

In an intriguing parallel in computational stylistics, scholars such as Paltridge, Whitelaw, Koppel, Argamon, Santini, and others have for similar reasons called for a *linguistic turn* in that subdiscipline.⁹⁵ Argamon's assessment regarding the centrality of linguistics for computational stylistics echoes that of Silva, Black, Louw, Nida, Levinsohn, and Porter in regards to the GNT. Argamon's perspective is worth quoting here at length:

Most work on computational stylistics to date has been based on hand-selected sets of content-independent features such as function words...[and]...Some interesting and effective feature sets have been found in this way...[and] function words have also proven to be surprisingly effective on their own. *In the long term, however, a clear foundation in a linguistic theory of meaning will be needed to gain true insight into the nature of the stylistic dimension(s) under study.*⁹⁶

These two separate linguistic turns, one in NT scholarship and one among researchers in computational stylistics, have patent implications for those pursuing computational stylistics in the GNT. Two contributing factors lie behind the separate "linguistic turns" in NT studies and in computational stylistics: the marked limitations of the traditional grammar framework, and the greater adequacy of certain functional schools of linguistics in explaining real-world language in use. I will treat each in turn.

Since Barr, a still small but growing minority of NT scholars (including but not limited to Silva, Black, Louw, Nida, Levinsohn, Cotterell, Turner, Porter, Fanning, McKay, and Reed) have built on aspects of Barr's program or initiated programs of their own. Silva concerned himself with reconstructing lexical semantics after Barr. See Silva, *Biblical Words and Their Meaning*. Cotterell and Turner's work (Cotterell and Turner, *Linguistics & Biblical Interpretation*) provided an introduction to basic concepts of linguistics and then focused on the utility of units of different rank to facilitate biblical interpretation. Porter and Fanning argued in substantially different ways for the centrality of aspect as the main semantic function of the Greek verbal system. See Porter, *Verbal Aspect in the Greek of the New Testament*; Fanning, *Verbal Aspect in New Testament Greek*. Black's *Linguistics for Students of New Testament Greek* is a primer introducing the classical subdisciplines of linguistics with NT examples, while his edited volume focuses on discourse analysis. Black, *Linguistics for Students of New Testament Greek*; Black et al., *Linguistics and New Testament Interpretation*. Last, Porter and Pitts have provided a terse review of developments in NT linguistics since Barr. Porter and Pitts, "New Testament Greek Language and Linguistics in Recent Research," *passim*.

⁹⁵ See Paltridge, "Genre Analysis and the Identification of Textual Boundaries," 288–97; Whitelaw et al., "Identifying Interpersonal Distance," 199; Koppel et al., "Computational Methods in Authorship Attribution," 12; Argamon and Koppel, "Finding Meaning in Stylistic Variation," 81; Santini, "Riding the Rough Waves of Genre on the Web," 3; Nini and Grant, "Bridging the Gap between Stylistic and Cognitive Approaches to Authorship," 177–82.

⁹⁶ Argamon and Koppel, "Finding Meaning in Stylistic Variation," 81.

2.5.3 Why Not Just Use Measures from Traditional Grammar?

Traditional grammar (TG)⁹⁷ is the familiar perspective that views language largely in terms of “parts of speech” (POS) categories and their inflections.⁹⁸ As a language construct, traditional grammar, to be sure, is peerless in terms of historical pedigree. It was originated by the early (though not earliest) Greek grammarians,⁹⁹ implemented in the Latin grammars of antiquity,¹⁰⁰ adopted by the classicist scholars of the Renaissance, replicated in the seminal Greek grammars of Beza and the Port Royal grammarians,¹⁰¹ and still remains (at least in grammars of the Koine) the predominant method of language analysis and pedagogy to this day.¹⁰² With such a pedigree, why consider any other approaches at all? The answer reduces to a single compelling reason. There now exists ample theoretical and applied *linguistic evidence* that the actual structure of the Koine, and indeed likely all languages, encodes far more types of information in far more categories than is currently captured by traditional grammar.

2.5.4 Why Use Linguistics Instead?

Four separate lines of evidence help to clarify the ways in which modern linguistics provides a more adequate account of language than does TG. The first of these relates to the

⁹⁷ See Lyons’ well-researched history of the development of traditional grammar (Lyons, *Introduction to Theoretical Linguistics*, 4–20).

⁹⁸ Lyons, *Introduction to Theoretical Linguistics*, 174. The “part of speech with inflections” perspective was recognized by the Stoics and codified by the Alexandrian philosophers as far back as the second century before the common era. See Lyons, *Introduction to Theoretical Linguistics*, 12.

⁹⁹ Interestingly, the part of speech construct of the earliest Stoic grammar was a functional one, rather close to the contemporary SFL category of theme and rheme. Only later was the construct of μέροι λόγου from the Τέκνη, Γραμματική fully replaced by word classes. Pitts helpfully locates the contributions of the earliest Greek grammarians in this regard; Pitts, “Greek Case in the Hellenistic and Byzantine Grammarians,” 262-73. Cf. also Halliday and Matthiessen, *An Introduction to Functional Grammar*, 51.

¹⁰⁰ For a review of the development of Latin grammar see Robins, *Diversions of Bloomsbury*, 185–203.

¹⁰¹ In one of the most laudatory effects of the digital age, Google, via its worldwide digitization project, has made accessible virtually all the Greek grammars from the sixteenth century onwards as free e-books.

¹⁰² While most grammars teach the Koine using traditional grammar, several have made notable attempts at Koine pedagogy from a linguistic perspective. Van Ness Goetchius’ little known grammar (Goetchius, *The Language of the New Testament*) was an early principled attempt to do so, followed more recently by Mounce’s very popular but rather linguistically uneven attempt to respect linguistic categories (Mounce, *Basics of Biblical Greek Grammar*). Porter, Reed and O’Donnell’s grammar is the most advanced current attempt to place New Testament Greek language pedagogy squarely upon a linguistic foundation; Porter et al., *Fundamentals of New Testament Greek*.

qualitative categorical richness of modern linguistic systems vs. TG, the second to the quantitative systemic potential of modern linguistic systems vs. TG, the third to the evolving utility of modern linguistics regarding the Koine of the GNT and the fourth to recently quantified categorical richness of modern linguistic systems regarding the Koine of the GNT.

Qualitative Categorical Richness of Modern Linguistic Systems vs. TG. Whether one initially favors or opposes modern linguistic theory, the last hundred years of theoretical and applied linguistic work since Saussure have been enormously productive in offering numerous testable theories of language. Setting aside for the moment some of the admittedly failed constructs of early twentieth century linguistic structuralism,¹⁰³ whether one interrogates, for instance, the tagmemics of Pike,¹⁰⁴ the linguistic analysis of the various functional schools, or the stratificational grammar of Sydney Lamb,¹⁰⁵ each proposes far more linguistic categories than TG. Moreover, these linguistic categories are largely novel; they are neither captured within the POS/inflectional framework of TG or else they require extensive qualifications to TG's categories to create such an equivalence. (See the discussion after Table 2.3 in this section.) We are not here arguing that having more and different categories makes for better categories.¹⁰⁶ Rather we are asserting that modern linguistics attempts to yield some sort of correspondence

¹⁰³ The early structuralists who followed Saussure, beginning with Franz Boas but especially Leonard Bloomfield and his followers, attempted to solidly establish linguistics as a *science*. In so doing, however, they veered into positivism and attached linguistics to what was then perceived as the halcyon achievement of human science in the early twentieth century: behaviorism. The heavily overloaded descriptivist program was demolished by Chomsky (Chomsky, *Syntactic Structures*, 17–25). For comparative overviews of the various linguistic schools and their historical trajectories see Davis, *Modern Theories of Language*, 14–336, and the more accessible treatment by Sampson, *Schools of Linguistics*, 34–235.

¹⁰⁴ Despite proposals by Pike and others regarding the generalizability of tagmemics to other disciplines, the remarkable achievement of the tagmeme (Pike's fundamental unit of language) lies principally in the almost algorithmic terseness of its four features: slot (where it can appear), class (what kind of unit it is), role (how it functions), and cohesion (how it relates to other units). These features efficiently recapitulate much of what is found in modern functional analysis. See Pike, *Linguistic Concepts*; Pike, *Language in Relation to a Unified Theory*.

¹⁰⁵ See Lamb and Newell, *An Outline of Stratificational Grammar*, 8–57.

¹⁰⁶ While a higher number of categories does, *ceteris paribus*, translate to a higher linguistic potential, it clearly does not necessitate in every case that SFL measures will outperform TG measures. Cf. Koppel et al., "Computational Methods in Authorship Attribution," 12–14.

between language *as it is actually used* and its description. That is, whether that correspondence is through the rather problematic competence/performance paradigm of the American Generativist schools following Chomsky or the stratal/systemic/functional paradigm of the functionalist/systemicist schools,¹⁰⁷ modern linguistic approaches attempt to tether their descriptions to the phenomena of language. This is what makes them both modern and linguistic. Leaving aside for the time being the important discussion of the relative utility of the Chomskyan competence/performance vs. the functionalist/language-in-use debate,¹⁰⁸ and its relation to the traditional praxis of science,¹⁰⁹ for the purpose of this primer all that needs to be demonstrated is the richer set of linguistic categories provided by either school as compared to traditional grammar. To do so in at least a notional way, I have enlisted the categories of Systemic Functional Linguistics (SFL).¹¹⁰ While justifying the full set of advantages behind this

¹⁰⁷ Modern linguistics has followed two very different trajectories with respect to correspondence since the landmark 1957 work by Chomsky, *Syntactic Structures*. The Generativist schools assert a competence/performance strategy for linguistic correspondence. This notion uses what others in the sciences would call “back-testing”—a valid method of verification indeed. The functional schools’ correspondence is structural and systemic, built up during the process of language analysis.

¹⁰⁸ In contrast to the Chomskyan tradition, SFL (which we will introduce shortly) is nonpropositional regarding universals, inductivist not introspectionist with respect to language phenomena, classificationalist not transformative with respect to method, and, in our view, statistically verificationist in its most mature formulations.

¹⁰⁹ Unfortunately, Chomskyan linguistics constitutes a remarkable epistemological narrowing of what scientific endeavor has historically meant by *correspondence*. Moreover, Chomskyan linguistics is curiously and anomalously anti-empirical; its priors require embracing what has not been established (language universals), its method narrows the epistemological boundary from the data of language to introspection and its verificationist method has narrowed from explanatory power to a subjective estimate of the correctness of invented sentences – and that by an indeterminate number of arbiters. Such a system leaves us little room to conclude anything other than that Chomskyan linguistics is indeed scientific in both its *ethos* and terminology but not in its *praxis*.

¹¹⁰ Halliday’s Systemic Functional Linguistics (SFL) is itself an outgrowth of Firth’s linguistics. Perhaps no better summary definition of SFL can be had than Halliday’s own description: “Systemic grammar is an analysis-synthesis grammar based on the paradigmatic notion of choice. It is built on the work of Saussure, Malinowski and Firth, Hjelmslev, the Prague school, and the American anthropological linguists Boas, Sapir, and Whorf; the main inspiration being J. R. Firth. It is a tristratal construct of semantics (meaning), lexicogrammar (wording), and phonology (sound). The organizing concept at each stratum is the paradigmatic system: A system is a set of options with an entry condition, such that exactly one option must be chosen if the entry condition is satisfied...the grammar as a whole is motivated with respect to the semantics. The only line of (relative) arbitrariness is that between content and expression (between the lexicogrammar and the phonology).” See Halliday, *On Grammar*, 262. Two particularly helpful introductions to SFL are Thompson, *Introducing Functional Grammar* and Teich, *Systemic Functional Grammar in Natural Language Generation*, 7–50.

choice of SFL,¹¹¹ as well as its limitations,¹¹² would take us too far from our current course, two factors relevant to the praxis of statistical linguistics can be stressed here. First, SFL explicitly formalizes what to a greater or lesser extent most linguists grant: that language encodes its sematic content at different ranks¹¹³ or spans of linguistic structure.¹¹⁴ Moreover, functional linguists have observed that each given rank in language not only serves a function at its own level of structure but functions as a “slot” within a larger structure of rank. While the *comprehensiveness* of the rank construct is under debate even among those who class themselves as systemic functionalists, what is clear is that language operates constitutively and simultaneously at *multiple ranks* to construe meaning. Accordingly, it must be so measured if we are ever to get our arms around what is summatively encoded in language. Second, Firth’s work, especially as it was later embodied by Halliday’s Scale and Category Grammar¹¹⁵ which subsequently evolved into SFL, recognizes that language categories must derive from “language-in-use,”¹¹⁶ language in terms of the natural systems of choices made by real speakers/writers of

¹¹¹ While Halliday’s own introduction (Halliday and Matthiessen, *An Introduction to Functional Grammar*) delineates many such advantages, it does little to explain his feel for language-in-use, nor how amenable his system is to quantification. For these see Halliday, *On Grammar*, and Halliday, *Computational and Quantitative Studies*. SFL welcomes empirical analysis, a perspective not shared by the various transformative (Chomskyan) approaches in particular.

¹¹² If traditional grammar can be faulted for its preoccupations with the rank of the word, SFL might be faulted for its paucity in developing consensus theories of language much beyond the clause. I am hardly the first to observe that language may be one of the disciplines in human endeavor we use most readily and understand least fully.

¹¹³ Rank, a term coined by Halliday, holds a central role in SFL because all systems (choice networks) in language originate at a given rank in the language (clause, word group, etc.). See especially Halliday and Matthiessen, *An Introduction to Functional Grammar*, 9–10. Most functional linguists agree that there exists some sort of scale of rank, perhaps even constituency rank in language, but agreement on the levels of that scale are often far from complete. For our purposes we will use Halliday’s rank scale (morpheme, word, word group, clause) but will omit morphemes from our initial explorations and add clause complexes. Cf. Halliday, *Linguistic Studies of Text and Discourse*, 24.

¹¹⁴ We use structure here in the non-linguistic sense of scaffolding. In SFL parlance *structure* relates to the syntagmatic hierarchy of language (its rank) and is a counterpoint to the paradigmatic *system*. Cf. Halliday and Matthiessen, *An Introduction to Functional Grammar*, 20.

¹¹⁵ Halliday, “Categories of the Theory of Grammar,” 241–92

¹¹⁶ The “language-in-use” perspective of SFL is most clearly codified via its system diagrams which are essentially maps of the systemic choices speakers or writers make when they communicate. This perspective is a broad and growing phenomenon in the linguistic landscape per Griffiths et al., *Language in Use*.

that language. Because the categories of SFL come “up” from the language itself,¹¹⁷ this naturally intersects with the praxis of modern statistical and corpus linguistics¹¹⁸ to operate upon language *in situ*—language as it exists in real corpora. It is here where we must be clear. The advantage of the functional schools in general and SFL in particular is that its categories actually *do* exist. The categories of an SFL-analyzed language were not produced introspectively by someone thinking great thoughts about language! Rather, they were mapped out by linguists in the field observing and analyzing the choices made by real speakers and real writers. These choices were then translated into system diagrams, where, depending on your perspective, Halliday has either freighted them with obscurantist terminology or given them names that are crystalline in their clarity. With this as background, I have organized below a sampling of high-level functional categories from SFL organized by rank (*in bold italics*) and compared them to the main paradigmatic categories of TG (underlined). In this notional example the target language was English. I have, however, informally inspected the GNT and confirmed that even at this coarse level of detail (delicacy), Table 2.3 would have been substantially identical had it been prepared from the Koine of the GNT. Where needed I have included the linguistic terminology in parentheses.¹¹⁹

¹¹⁷ Halliday’s wry comment might be recalled here, “It has always seemed to me that grammar was a subject with too much theory and too little data.” Halliday, *Computational and Quantitative Studies*, 76.

¹¹⁸ Corpus linguistics, according to Biber’s helpful definition, is the empirical study of large and principled collections of natural texts using both qualitative and quantitative analysis techniques (Biber et al., *Corpus Linguistics*, 4). For the utility of corpus linguistics applied to the GNT see O’Donnell, *Corpus Linguistics*.

¹¹⁹ Where there exists, no widespread agreement on linguistic terminology, I have defaulted to SFL terminology.

Categories of Linguistic Rank Absent or Inefficiently Captured by Traditional Grammar
(Such Categories Are Indicated by ***Bold Italics***.)¹²⁰

<i>Size of Unit (Rank)</i>	<i>Construct Missing or Largely Missing</i>	<i>Linguistic Definition</i>	<i>Example Language Measures</i>
<i>Word (Lexis)</i>	Vocabulary breadth	The base form of a lexeme	<i>Base form of a word (lemmas or base forms of the lexeme)</i> ¹²¹
	Vocabulary depth	Inflected lexeme	<i>Identical forms of words (inflected lexemes)</i>
	Lexical semantic breadth	Primary lexical semantic categories	<i>The basic categories of word meanings (semantic domains)</i> ¹²²
	Lexical semantic depth ¹²³	Extended lexical semantic categories	<i>More detailed (delicate) categories of word meanings (semantic subdomains)</i>
<i>Phrase (Word Group)</i>	What is happening?	Experiential (meta)function	<i>Determiners (deixis)</i> ¹²⁴ , <i>classifiers, qualifiers, action perspective (aspect) number, degree, case</i>
	Who is involved? (participants)	Interpersonal (meta)function	<i>Person, negation (polarity), judgment (appraisal), words (modal comment, etc.) mood, gender</i>
	What helps get the idea across?	Textual (meta)function	<i>Voice, conjunctions, continuatives</i>
<i>Clause</i>	What is happening?	Experiential (meta)function	<i>Types of actions (process types), participant types, circumstance types, tense, interrog. types, equatives/ergatives (transitivity)</i>
	Who is involved? (participant roles)	Interpersonal (meta)function	<i>Subject, direct object (complement), judgment /attitude phrases (modal comment adjuncts, and other adjunct classes), direct address (vocative)</i>
	What helps get the idea across?	Textual (meta)function	<i>Listener/speaker orientation (theme-rheme vs. given-new)</i>
<i>Clause Chains (Clause Complex)</i>	What is happening?	Experiential (meta)function	<i>Types of relation (temporal, spatial, etc.), types of complexes (expansion: elaborating, extending, enhancing; projection: ideas, reports, facts)</i>
	Logical relations	Logical (meta)function	<i>Coordination, subordination (taxis), (conjunctive adjuncts)</i>
<i>Structures That Cross Rank Boundaries</i>	Lexical priming ¹²⁵	Word interactions that yield semantics	<i>Word patterns (collocations), word or clausal patterns (colligations)</i>
	Cohesion	Patterns causing texts to make sense	<i>Reference, substitution, ellipsis, lexical cohesion, etc.</i> ¹²⁶
	“Word” order	Slot order	<i>Order (permutation) of word, clause, and clause complex group slots.</i>

Table 2.3

Several observations aid in the interpretation of Table 2.3. First, if the categories in bold italics actually exist—and they do because SFL derives them from actual language instances—

¹²⁰ Though these categories are only of rank, our categories of rank derive from a systemic functional perspective; compare their similarity with that of Mollet et al., “Choosing the Best Tools,” 437–62.

¹²¹ Lemmas are found in all modern traditional grammatical annotations of the GNT, however.

¹²² For the theory of the semantic domains related to the GNT see Louw, *Semantics of New Testament Greek*.

¹²³ Many semantic measures exist (e.g. Zipf’s delta) but cannot be explored here; Zipf, *Selected Studies of the Principle of Relative Frequency*, 8–30.

¹²⁴ For this and all other linguistic terminology in Table 2.3 (except lexical priming), see Halliday and Matthiessen, *An Introduction to Functional Grammar*.

¹²⁵ Hoey, *Lexical Priming*, 16–113.

¹²⁶ Halliday and Hasan, *Cohesion in English*, 31–292.

then the table demonstrates that traditional grammar *is* indeed missing large swaths of “things” in language.¹²⁷ Moreover, similar charts could be prepared for the *paradigmatic system* (the detail or delicacy of the horizontal axis of choices) as well as metafunctions of language. By any measure then, traditional grammar is, in fact, a very limited description of language. Second, SFL largely incorporates TG’s inflectional categories (i.e. a verb’s tense, voice, mood, person, and number) within its paradigmatic systems. Given this, in a way quite similar to other disciplines, as an historically posterior systematic, SFL can be seen to be a generalization of the historically prior one, traditional grammar.¹²⁸ Third, the table is merely illustrative, not exhaustive. It exemplifies the kinds of things that are missing in traditional grammar. An actual quantification of the number of missing categories can be developed by separately calculating the *systemic potential* of any two theories and inspecting their overlap. This exercise is presented below. Fourth, one might quite reasonably object that by simply adding Boolean logic (if-then-else rules) to traditional grammar the missing linguistic categories in Table 2.3 can be recapitulated easily enough. This is quite true. Yet, this underscores our point: the POS/inflectional formalism of traditional grammar is *missing* those categories, necessitating

¹²⁷ Some might well argue that Halliday’s system networks were developed by the analysis of only two languages, Mandarin and English, and that only the system network of English was comprehensively worked out. Accordingly, SFL’s categories are vacuous for the Koine, especially given SFL’s resistance to language universals. This, in fact, misunderstands SFL. At the most coarse level of language description, SFL asserts that all languages do at least three things: they make sense of experience (the experiential metafunction), enact social relations (the interpersonal metafunction), and weave the two together using the assets of language (the textual metafunction). Halliday and Matthiessen, *An Introduction to Functional Grammar*, 29. For even the least delicate language with the fewest number of systems, SFL will yield many magnitudes of order more measures than that provided by TG. Moreover, the initial linguistic probes used in this work will operate only at this most general level of delicacy—the level that has the highest level of language universality.

¹²⁸ There is ample historical precedent for a prior systematic to be subsumed inside a larger one. This is one of the hallmarks of a maturing science. In this way, for instance, the mathematics of classical Newtonian physics was discovered to be a special case of the mathematics of relativity in unaccelerated reference frames. Three other hallmarks of a maturing science exist in my view: (1) a descriptivist program is followed by verification/prediction program, (2) complexity of the original descriptivist program can be pruned down to fundamental statements and (3) the science is able to be generalized to solve problems within other disciplines. In linguistics, SFL is arguably an instance of pruning, and Tagmemics is an early attempt at generalizability. For an application of the latter to rhetoric and style, see Young et al., *Rhetoric: Discovery and Change*.

feats of Boolean legerdemain to recover from the phenomena of language categories absent in TG.¹²⁹

Quantitative Systemic Potential of Modern Linguistic Systems vs. TG. Secondly, a *quantitative* case can be made for including modern linguistic categories in our work. To make that case, it is necessary to again turn to the systems of paradigmatic choices defined by Halliday. In SFL, as in any system network, the total number of possible selection expressions (the features in the system network that have been selected in the instance of language in question)¹³⁰ can be quite accurately approximated as the product of all possible (i.e. grammatical) choices in the entire system network of the language, with the simplifying assumption that all such choices are binary. Accordingly, the number of possibilities (the systemic potential) of even a relatively simple language like English, which has, say, 23 systems, exceeds over 70,000 categories.¹³¹ Languages with more and deeper systems contain *millions* of selection expressions and the most complex languages may contain *billions*.¹³² In contrast, traditional grammar with its POS paradigm applied to a formal definition of the NT Koine encodes, by our count, 1,635 combinations of POS categories and their inflections at the rank of

¹²⁹ This goes far to mitigate Ledger's assumption that paradigmatically delicate languages (in terms of highly inflected POS) will yield more readily to decompositional multivariate analytics. Given the strikingly more delicate number of linguistic species available with SFL, however, this perceived advantage of highly inflected languages recedes to negligibility. See Ledger, "A New Approach to Stylometry," 71.

¹³⁰ Halliday, *Computational and Quantitative Studies*, 257.

¹³¹ These numbers come from Halliday's system network of English (Halliday, *On Grammar*, 393). More detail on this (as well as a larger estimate) is provided in Section 4.3.4.

¹³² Halliday further explains that some languages have over forty systems. Assuming that the choices are binary opposition, these languages can produce over 1 trillion selection expressions (Halliday, *On Grammar*, 392). The magnitude of these numbers has another surprising ramification: the richness of the choices provided by the language system far exceeds the working vocabulary of the average person operating in that same language. When, as it often does, system potential (delicacy) exceeds working lexical capacity we begin to appreciate the relative importance of the paradigmatic cline of language, an appreciation that must be reflected in our linguistic description and analysis.

lexis.¹³³ Aside from its face validity, the advantage of a linguistic formalism that yields a higher number of distinct measures is at least twofold. First, as the number of truly distinct measures used by a researcher expands, one necessarily more closely approximates the linguistic *potential* of the language. Second, and for the same reason, empirical work has demonstrated that the highest accuracies in applied computational stylistic tasks have been achieved as the categories and the species within each category become larger.¹³⁴

Evolving Utility of Modern Linguistics Regarding the Koine of the GNT: Third, the large number of categories achievable by modern linguistic analysis and annotation no longer lies in the realm of theory alone when it comes to the Greek of the New Testament. As we have already alluded to, close to two hundred articles, monographs, dissertations, and books detailing the application of modern linguistic theory to the Bible have appeared since Barr made his groundbreaking argument for modern language theory in the service of biblical semantics in 1961.¹³⁵ A meta-analysis of this literature demonstrates that much of this work operates at or above the level of the clause, precisely at a level (or rank) at which TG historically does not operate.¹³⁶

¹³³ If TG's functional categories are permitted (which requires information to be borrowed from surrounding words to disambiguate the function of the word in question) this number will increase by approximately a magnitude of order.

¹³⁴ Tambouratzis, "Assessing the Effectiveness of Feature Groups," 249.

¹³⁵ A partial list of the more important works is included in the 298-item bibliography in Porter and Pitts, "New Testament Greek Language and Linguistics in Recent Research," 241–55. By our count 178 of these relate directly to biblical linguistics. While it must be granted that some of this literature is uneven, sometimes lacking theoretical grounding and quite often lacking any sort of verificational method, such is often the case when an academic discipline is emerging from its adolescence.

¹³⁶ Although Dyscolus had a volume on syntax (see Householder, *The Syntax of Apollonius Dyscolus*) and the Port Royal Grammarians liberally borrowed from Arnaud's *Grammaire Générale* for their insights on syntax, for most of history syntactic categories were barely distinguishable from logical categories (taxis and the like). Yet, beginning in the nineteenth century, language study began to be more fruitfully expanded beyond its traditional domain (lexis) under the influence of what would become a remarkable intellectual cascade inaugurated by the early nineteenth century comparative philologists. See especially the foundational works by von Schlegel, *Ueber die Sprache und Weisheit der Indier* and Humboldt, *Ueber die Kawi-Sprache auf der Insel Java*, followed by Grimm, *Deutsche Grammatik*; Grimm, *Geschichte der Deutschen Sprache* and Bopp, *A Comparative Grammar*. The era of philology reached its apex under Delbrück. See especially Delbrück, *Introduction to the Study of Language* and Brugmann, *A Comparative Grammar of the Indo-Germanic Languages*. When Ferrar (Ferrar, *A Comparative Grammar of Sanskrit*) and others brought the work of the comparative philologists into a study of traditional Greek grammar, and Clyde published his widely influential work on Greek Syntax (Clyde, *Greek Syntax, with a Rationale of the*

Recently Quantified Categorical Richness of Modern Linguistic Systems Regarding the Koine of the GNT. Fourth, our own research in the Koine has demonstrated that (a) empirically derived categories of the GNT display extraordinary delicacy (i.e. a remarkable number of categories) especially among units of linguistic rank above lexis,¹³⁷ and (b) many of these categories display statistically significant differences between genre and authorship theories. Specifically, GNT clauses alone exist in 1,412 categories. Moreover, just one of the varieties of clause complexes we have induced from the eclectic text of the GNT (complexes annotated with only the interpersonal functional notation from OpenText)¹³⁸ exists in 12,303 categories. While doubtless many of these categories will prove to be statistically unuseful, this bounty underscores

Constructions) standard Greek grammars then began to incorporate syntax as a pedagogical category as a matter of course. Cf. Jelf, *A Grammar of the Greek Language: Volume II: Syntax*; Crosby, *A Compendious Grammar of the Greek Language*; Winer, *A Grammar of the Idioms of the Greek Language of the New Testament*. While the era of comparative philology/grammar has largely given way to modern linguistics, and while their nineteenth-century methods of diachronic inference have largely yielded to more quantitative and verifiable synchronic methods of language comparison, what has now become known as comparative linguistics is still quite capable of formulating colligative statements of language universals, including, mostly notably, Greenberg's seminal work. Greenberg et al., eds., *Universals of Human Language*. How shall we evaluate this broad a trajectory? First, by asking the broader questions of similarity, connection, and language evolution, comparative philology not only became the intellectual precursor to modern linguistics but still influences modern comparative linguistics today. Works like Greenberg's identify connections that are helpful in their own right, even if we are to conclude that many of his universals are cognitively or physiologically determined rather than necessarily reflective of language phylogeny. Second, despite the Herculean labors of the comparative philologists, on balance their effect on the conception of syntax by traditional grammarians has been less than salutary. Precisely because traditional grammar has adopted the terminology and methodology of the comparativists, their categorizations display remarkable classificational shallowness – a necessary outcome of an exercise whose sole purpose was *cross-language* analysis, rather than *within-language* analysis. The limited classificational depth of TG compared to the much higher categorical potential within the paradigmatic systems of SFL is, in fact, even more marked than those found in the syntagmatic dimension of language. Second, their classifications are often *ex post facto* linguistically in that they impress upon lower linguistic units of rank information (frequently logical) from the next higher unit in rank. This exogenous borrowing of content obfuscates the functional contribution of the rank in question. Modern and quite popular examples of syntax based on traditional grammar have done little to overcome these deficits (e.g. Wallace, *Greek Grammar Beyond the Basics*; Wallace, *The Basics of New Testament Syntax*). A similar conclusion can be issued concerning the electronically published Lexham Syntactic databases of Logos Bible software (Lukaszewski, *The Lexham Syntactic Greek New Testament*).

¹³⁷ This is especially true using the SFL clause annotation of the OpenText GNT above the rank of the clause. See Section 5.1.

¹³⁸ Stanley E. Porter, the originator of the OpenText.org's clausal annotation model, helpfully clarifies that at the time when the clausal annotation of OpenText was being developed it was based on the interpersonal/ideational metafunction. Halliday subsequently split the ideational from the interpersonal metafunctions in the fourth edition of his *Introduction to Systemic Functional Grammar*. (Porter, personal communication.)

that the corpus itself preserves in its record a rich repository of actualized choices from the author/editor/redactor. These too are absent from the formalism of traditional grammar.

In summary, the best modern linguistic evidence in the last century demonstrates that virtually all languages analyzed to date are *multistructural*, *multisystemic*, and *multifunctional*. This means that languages express their syntagmatic content in multiple levels of rank, their paradigmatic content by, at the very least, dozens of systems with dozens of binary or multinomial choices, and that these choices summatively achieve (realize) multiple communicative functions simultaneously. By any measure, then, modern linguistics, exemplified here by SFL, provides many more “things to test” than traditional grammar. These summary findings about language necessitate that we do *not* limit our description and analysis of the Koine to traditional grammar.

2.6 Before Beginning: Haven't We Forgotten Something?

Following Table 2.1, I have now defined our starting stances (priors), explored the secondary problems to be solved (the “gaps” in the GNT), and the tools needed (linguistics and computational stylistics) to address those gaps. Accordingly it is now time to explore whether there are potential solutions that might already exist to aid in addressing both our primary problem (the lack of a theory) and our secondary problem (the gaps in GNT computational stylistics). Two areas might be suggested. These will be presented in subsections 2.6.1 and sections 2.6.2.

2.6.1 Closing Some Philosophical and Hermeneutical Gaps

Recall from Section 2.3.1 that the first critique that might be leveled against computational stylistics is the charge of a lurking positivism hidden underneath a remonstrant empiricism. While the equation of empirico-positivism is to be rejected at face value—for not all empirical methods are positivistic—this critique nonetheless hits home in terms of the historic praxis of stylometry. Early to mid-twentieth-century NT stylometry was, in essence, an uninhibited positivist bacchanal. What might mitigate such a charge in this current work? My response is that abduction itself paired with an experimental design sufficient to mitigate Type I and Type II error all operating under a solely coherentist framework should be sufficient to defang the charge of positivism. And what of the remaining four critiques? They are naïve empiricism, discursive failure to triangulate findings with the long history of accepted diachronic praxis in NT studies, failure to appropriate post-positivistic (mixed method) or multiple methods of verification to results, and the failure to test or consider necessary truths in the design. In each of these, mid- to late-twentieth-century stylometry is guilty as charged. How, then, can a praxis

be articulated to relieve the current effort of these same charges? While the synthetic proposal in Chapter Four is explicitly intended to do exactly this, two clarifying points should be made here.

First, the force of this combined critique must be felt by NT scholars in general, not just NT scholars of the quantitative stripe. That is, this charge is clearly not limited to nor primarily directed at stylometric praxis alone. It is to be granted, then, that though it is philosophically dead, a broad undercurrent of retrenched positivism still courses freely throughout New Testament studies today. Per Sherwood and Moore, this undercurrent not only proposed and implemented biased criteria and methods during the heyday of historical criticism,¹³⁹ it also served to accelerate the fragmentation, methodological dissonance, and scholarly insularity characteristic of the postmodern turn in the social and cultural sciences.

Our assertion is that an abductive approach again comes to our aid for many of these issues. Specifically, our abductive approach employs the following:

1. *A philosophically informed reentrant approach to NT research* (including articulating its priors). This mitigates charges of foundationalist priors of naïve empiricism. It does this two ways. First, abductive priors are not foundationalist in that (a) any testable priors *are permitted* via abductive inference and most importantly and (b) any prior may be jettisoned in the process. A foundationalist prior that gets jettisoned can hardly be called foundationalist. This is a coherentist exercise through and through.
2. *A mixed methodology data collection process* that collects the relevant qualitative data from the entire history of computation stylistics (Chapter Three). The design in Chapter Four, moreover, will present some tools and methods by which abductive findings can be triangulated to the largely diachronic praxis of NT studies.

¹³⁹ Sherwood and Moore piercingly characterize this as the period of the sub-sub-sub specialist in the period of the Enlightenment Bible (Moore and Sherwood, "Biblical Studies 'After' Theory: (Part3)," 192–94).

3. *A verificational model* that embraces multiple methods of verification, including classical, Bayesian and mixed methods research. This directly addresses the critiques found in section 2.2 and 2.3.

2.6.2 Closing Some Experimental Design Gaps

2.6.2.1 The Multiplication of Type I and Type II Error *Both Statistically and Causally*

As we have just seen, the process of abduction aids in the reduction of Type I and Type II error through its use of multiple modes of verification and its own intrinsic iterative nature. What is little recognized, however, is that both types of errors can occur at two distinct points in a research study: at the point of the statistical hypothesis test used to determine statistical significance *and* at the point of interpretation when that statistical significance is assigned a cause. We will first distinguish between these two modes of errors and then describe how our approach seeks to mitigate these errors.

A Clarification of Error Type and Error Points in Computational Stylistics. A statistical false positive (Type I error) occurs when a given statistical test “says” something is significant (typically by rejecting the null hypothesis) when, in fact, that difference is not significant. A statistical false negative (Type II error) is the converse. A *causal* false positive (Causal Type I error) occurs when a truly significant difference for a given finding (say, between two sets of texts on a given linguistic measure) is claimed by the researcher to be due to a given cause but it is not due to that cause. A causal false negative (Causal Type II error) is the converse.

Complicating matters further, a significant difference may be due to the combined effects of one or more causes and their interactions. (An interaction is an effect that arises when the combination of two main effects is significantly greater or lesser together than their individual effects.) To make our complications complete, it is possible for a given finding to commit *both*

Type I and Type II errors simultaneously. In terms of causal assignments, this can occur when a conclusion is made that the cause of a significant finding is A and not B, and the cause of the significance is, in reality, B and not A (or alternatively, B *and* C).

An Abductive Approach to Mitigating Error Types and Modes. To address these multiple types and modes of error experiments have been built into the design (see Chapter Four) that (1) test significance both statistically and causally, (2) test for alternative causes,¹⁴⁰ (3) test for multiple causation, (4) test for interactions between possible causes,¹⁴¹ (5) test that different profiles of causation may apply to different language strata, paradigms, syntagms, and functions and (6) demonstrate directionality, that the effect is due to the cause and not the converse.¹⁴²

2.6.2.2 The Unsupervised/Supervised Antinomy

The Problem of Supervised and Unsupervised Approaches. How might we at this juncture address the antinomy discussed in Section 2.4.3.2.2? That is, how can we avoid the poorer classification produced by unsupervised methods on the one hand, and causal misassignment (due to the elimination of covariant structure) by supervised methods on the other? First, an abductive interplay will be built between unsupervised and supervised methods in our experimental design. That is, both methods will be executed and their results compared. Second, other experimental design approaches exist that use the covariance structure within the text rather than ignoring it. These include employing covariant structure models (e.g. SEM) that use rather than throw away covariant structure data. They, too, will be executed in subsequent abductive cycles defined in the design.

¹⁴⁰ Alternative causes can be tested by a number of strategies including treatment by blocks (Brown and Melamed, *Experimental Design and Analysis*, 36–43) and confirmatory factor analysis and other latent structural models.

¹⁴¹ Interactions can be tested via parametric designs such as MANOVA and ANCOVA (Bray, *Multivariate Analysis of Variance*, 13–54), and via decompositional nonparametric designs such as hierarchical log-linear analysis (Knoke and Burke, *Log-Linear Models*, 11–30).

¹⁴² For an extended discussion of the mathematical determination of directionality, see Davis, *The Logic of Causal Order*, 7–68.

2.7 Remind Me Again: Why Is This Theology?

In one last ground-clearing exercise I anticipate a question that some readers may have at this juncture. Given the extensive use of linguistics and mathematics in this work, what, exactly, makes this a work in *theology* rather than a work in *computational stylistics* merely applied to the GNT? Five distinct things, I think. *First*, this is a dissertation following a well-worn path, demonstrating openness to synchronic insights from the original language of the text, even if those insights are linguistically and mathematically construed. In our case, moreover, this includes openness to the possibility that the text encodes sociolectic information, clues to the context of situation and context of culture which we have not yet discovered. If indeed an underlying *latent structure* exists amidst the observed linguistic pluralities and diversities in the GNT, one that differentially encodes sociolectic or idiolectic information, such a thing would make a solid contribution to NT scholarship. *Second*, our starting point is thoroughly theological. We are addressing long-standing issues in NT scholarship, issues surrounding the intersections of hermeneutics and philosophy on the one hand and antinomies regarding authorship and genre on the other. Insights from this quarter, too, would have necessary implications on any scholarly reconstruction of early Christianity. *Third*, not only is the fundamental data of this work the language of the New Testament itself, the use of modern linguistics stands solidly in the synchronic text-centric traditions of Christian humanism that emerged in the Renaissance and flowered in the exegesis of the Reformers.¹⁴³ *Fourth*, as shall be shortly demonstrated, the six authorship hypotheses have been drawn directly from the history of New Testament scholarship itself. *Fifth*, as can be inspected from the last chapter, I have cast this issue in terms that I believe it must be cast, as an issue in *theological hermeneutics*. The question, in the final sense, is not

¹⁴³ In addition, this linguistically developed approach respects the so-called linguistic turn in New Testament studies.

whether empirical methods applied to the GNT have been, quite frequently, poorly employed. Assuredly they have! The larger question relates to the eventual role such methods might have in New Testament studies. What if, indeed, a nonpositivist, experimentally-controlled, causally-disciplined, linguistically-informed approach *did* produce consistent, validated findings? Such a thing would have necessary and even potentially wide-ranging implications on the current, energetic dialectic in theological hermeneutics. It would interrogate priors and praxis on both sides of the historic glass wall that has, over time, come to characterize the modern academy.¹⁴⁴ That, too, I think, would be a good thing.

¹⁴⁴ Everyone with more than a passing interest in theological hermeneutics recalls that the historic division between the arts and the sciences in the Western academy was in large part inaugurated and institutionalized by the Scholastics in the twelfth century. The decision then to create a bifurcated architecture of that historic event, produces now what was unanticipated then, a *Weltanschauung* of the arts and a *Weltanschauung* of the sciences—an academy divided—where we talk past one another and delegitimize one another.

3 History

3.1 What Is an Experimental Design Review?

An experimental design review, as implemented here, is quite distinct from a traditional historical review. Its objective is not to interpret researchers in the light of their times, *per se*, but rather in light of what their praxis tells *us* about experimental design. Said colloquially, an *experimental design review* asks three questions: (1) What did prior research get “wrong”?¹ (2) What did it get “right”? (3) *How* do those answers inform the central goal of *this* research—the disentangling of genre from authorship in the GNT? Thus, as defined here, an experimental design review must eventually yield (*ipso facto!*) an experimental *design*—a step-by-step approach to an inductive (or in our case an abductive) set of experiments. Toward that end four format differences exist between a standard historical review and an experimental design review. *First*, as with a traditional theological review, an experimental design can be historically structured, but we cannot limit ourselves to computational stylistic studies executed only upon the biblical texts. Doing so would cover only about ten percent of the total studies in computational stylistics. Moreover, many of the recent “best practices” have not emerged from within our discipline or have not yet been executed upon the biblical texts. (Recall the “gaps” in Section 2.3.) *Second*, this review has discovered that quite often researchers make the same mistakes across historical eras. Rather than disturb the flow of the review to point out when each design implication occurs, I will employ a relatively unobtrusive alphabetic superscript in square brackets (^[a]) to indicate experimental design implications. Superscripts will restart at each second-level section (e.g. 3.2, 3.3 etc.).² Those design implications will then be comprehensively

¹ By “wrong” is meant the determinations made in prior research that produced Type I or Type II error—either in terms of measurement, analysis, or in the assignment of causation.

² When a footnote is also present, the footnote will precede the design implication brackets.

collected in two summary experimental design tables in the Appendix. These, in turn, have been abbreviated into a single table in Chapter Four: (Table 4.1). Those tables will allow us to learn from past experimental omissions, leverage best practices, and allow us to inspect repeated themes at a glance. *Third*, in contrast to an historical review, generally, there will be no attempt to critically appraise any given scholar *per se*. More extended author-level evaluations (such as A.Q. Morton) will be executed only when they serve the larger purposes of furthering the development of the experimental design. In fact, while this conclusion can perhaps be contested, it seems far more frequently to be the case that the failures, misinterpretations and contradictions of the discipline—and there have been many—were not due to the wholesale incompetence of its researchers. Rather, an experimental design orientation demonstrates that many otherwise competent researchers were freighted with overly simplistic models of language, inference, and causality that held sway throughout most of the long history of stylometry/computational stylistics. The most recent thirty years of research has made the extent of this “freighting” abundantly clear. This alone underscores why an experimental design review is necessary. *Fourth*, this review is extensive because every historical period, in my view, is relevant to developing the experimental design summarized in Chapter Four. Accordingly, this review will briefly cover seven eras and five milestones in stylometry/computational stylistics. There is no doubt that other reviewers will see these periods somewhat differently in a number of minor and perhaps a few major ways. But how we might draw the lines is hardly the point. The examples of the past, whether for good or for ill, must inform our efforts. The resulting design will be all the more robust for having listened to those voices.

3.2 Period One: Quantitative Stylistics from Antiquity to Campbell

3.2.1 The Quantitative Description of Texts in Antiquity

The quantitative description of texts (a sort of meta-language to express the content, meaning and emphases within a single text or across a corpus) stretches back as far as texts themselves. If, for instance, the attitude of the early Sanskrit grammarians or the scholars of Alexandria can be considered normative, no dichotomy existed between the value vested in the qualitative and quantitative description of texts. Here Delcourt's summary is helpful:

The quantitative description of texts originated with the wish to preserve highly-valued works. The *Anuvâkânukramani*, which is traditionally attributed to the Sanskrit grammarian Saunaka...composed between 500 and 200 B.C., records that the Rgveda contains 1,017 hymns, 10,580.5 stanzas, 153,826 words and 432,000 syllables...The Alexandrian scholars, among whom the most prominent was Aristarchus of Samothrace (c. 217–145 B.C.), noted that such-and-such a word, sense or construction in the Iliad, in the Odyssey or in other Greek masterpieces ἀει κεχρήται (« is always used ») or constitutes a σπανίως λεγόμενον or εἰρημένον (« seldom [used] expression »), a ἀπαξ λεγόμενον or εἰρημένον («once [used] expression»), and so on... Finally, the Masorettes—whose early forerunners are clearly evoked by the Talmudic tractate יְשִׁבְּרָה (3rd-5th Century) ...provided the number of verses, of words, of letters, etc. the middle verse, the middle word, the middle letter, etc. of the Hebrew Bible, of its different books and of its subdivisions.³

To supplement Delcourt's point, we would only add that the intertwined integration—even to the point of mystic excess—between texts and their numeric description by Arabic, Jewish, Gnostic and Christian numerologists clearly underscores the high value ancients placed on quantification. Quantification, for them, was a parallel voice in understanding and interpreting texts.⁴

³ Delcourt, "Stylometry," 979–80.

⁴ Hopper, *Medieval Number Symbolism*, 50–135.

3.2.2 Augustine to the Enlightenment

Augustine continued the perspective that qualitative and quantitative descriptions were both valid and important for a comprehensive Christian hermeneutic. Although he clearly privileged a text-centric approach in his praxis, Augustine, too, viewed both exercises as parallel, valued resources in textual interpretation:

....we must not despise the science of numbers, which, in many passages of Holy Scripture, is found to be of eminent service to the careful interpreter. Neither has it been without reason numbered among God's praises, "Thou hast ordered all things in number, and measure, and weight."⁵

Elsewhere I have executed a study on the early reception history of quantitative methods in the broader academy and in the theological enterprise.⁶ Consistent with the findings from antiquity and Augustine's high view of the value of quantification just described, that study yielded three distinct conclusions: (1) Neither quantitative, nor qualitative, nor historical methods enjoyed an exclusive hegemony over any other method before the Scholastic period in the West. Each was interpretively valued. (2) This relative egalitarianism was disrupted by two seminal fragmentation events which privileged and isolated certain methods within separate verticals within the academy. The first of these fragmentation events was the division of the Western academy itself into two separate courses of study, the *Trivium* (the arts) and the *Quadrivium* (the sciences), which was institutionalized, though not begun, in the period of Western scholastics. The second fragmentation event, ironically, followed Gabler's attempt to unify the theological landscape under an optimistic, but ultimately unsuccessful, coherentist exegetical program. These dual fragmentation events predisposed the modern theological

⁵ Schaff, *St Augustin's The City of God*, 223.

⁶ Libby, "Locating Computational Methods within the Theological Spectrum," (forthcoming).

enterprise both by training and by choice to exclude serious and/or an informed consideration of quantitative methods.

3.3 Period Two: Stylometric⁷ Descriptivism (1867–1935)

3.3.1 Late Nineteenth-Century Stylometric Descriptivism (1867–1900)

Stylometry, according to many of its reviewers,⁸ emerged when Thomas Mendenhall took up Augustus de Morgan’s suggestion that word lengths could discriminate between authors and then published his now famous “curves of composition.”⁹ It seems far better, however, to trace the beginning of post-Enlightenment stylometry to three prior pioneers: Lewis Campbell, Wincenty Lutosławski, and Wilhelm Dittenberger.

Lewis Campbell, the great nineteenth-century Classical scholar at St. Andrews, is likely the best known of these four pioneers to students of theology. What is less known, likely, is that Campbell, in his massive work of classical erudition *Sophistes and Politicus of Plato*, became the first post-Enlightenment scholar to issue wide-ranging quantitative stylistic observations. Among these, Campbell noticed that (a) many words do not recur in Plato (according to the then-accepted chronology of Plato),¹⁰ (b) the corpus displayed peculiarities of diction and (c) there were 1,492 *hapax legomena*, words that only occur once, in *Timaeus*, *Critias* and *Laws*.¹¹

⁷ The origin of the term “stylometry” is seldom discussed. During the mid-nineteenth century, quantitative description in literary practice enjoyed a revival. A small group of scholars, among them Fleary, Ingram and Furnival, developed a technique they termed “Stylometrics” which involved measuring the variation of meter in poetry and lexical repetition in prose. The school reached its height, according to Tuldava, with the founding of the New Shakespearean Society in the 1870’s; Tuldava, “The Development of Statistical Stylistics,” 145–6. This revival was likely part of the broader expansion of quantitative methods within nineteenth century philology. This is almost certain, for Tuldava records that “Among the members of the Society were...Fleary...Ingram... [and] Furnival.” See Tuldava, “The Development of Statistical Stylistics,” 145. Cf. also Tuldava, “Stylistics, Author Identification,” 370–1.

⁸ So Campbell et al., “The Provenance of *De Doctrina Christiana*,” 104; Malyutov, “Authorship Attribution of Texts,” 363; Alviar, “Recent Advances in Computational Linguistics,” 140; Mendenhall, “A Mechanical Solution of a Literary Problem,” *passim*.

⁹ Mendenhall, “Characteristic Curves of Composition,” 237–49.

¹⁰ Campbell, *Sophistes and Politicus*, xxv–xxvii.

¹¹ Campbell, *Sophistes and Politicus*, xxxi.

Accordingly, fifty years prior to P.N. Harrison's famous argument, we find *hapax legomena*^[a] raised in relation to authorship. Yet—and this point seems often missed—Campbell associated word variety with *genre rather than authorship*.^{12[b]}

On the continent Dittenberger chose a different path than Campbell: the use of high-frequency words,^[c] characteristic word groups (e.g., *καὶ μὴν*), and synonyms^[d] to compare and contrast style.¹³ As Campbell before him and Lutosławski after him, Dittenberger's original focus was not upon authorship per se, as much as it was upon *stylochronometry*,^[e] the attempt to date works within an author's corpus. In comparing Campbell and Dittenberger, clear differences between their respective approaches can be observed; Dittenberger's higher frequency words were less susceptible to the accidental appearance or disappearance of a lexical form¹⁴ but it was based on far fewer words than Campbell's approach. Intriguingly these two distinct tendencies would wend their way through the next seventy years of stylometric praxis.

The third inaugural scholar of this review, the Polish author/philosopher Wincenty Lutosławski, published two major works on stylometry, *Principes de Stylométrie* and *The Origin and Growth of Plato's Logic*.¹⁵ In an odd injustice of history, this scholar is relatively unknown today to the more casual students of literary stylistics—in perhaps the same way Campbell was initially unknown to continental stylometricians.¹⁶ Lutosławski, however, in both his major

¹² Hence, for Campbell, the dialectical *Sophist* is *expected* to have a fewer variety of words than the *Republic*, just as *Timaeus*, *Critias* and *Laws* are fully expected to have many more *hapax legomena* “due to peculiarity of subject.” Campbell, *Sophistes and Politicus*, xxxi-xxxiii. Greenspahn, after analyzing *hapax* in the *Tanakh* concluded, “Poetry uses more *hapax legomena* than prose... Concentrations of rare forms in general and *hapax legomena* in particular, although not a useful test of authorship, is nonetheless a characteristic of style.” Greenspahn, “The Number and Distribution of Hapax Legomena,” 18–19.

¹³ Dittenberger, “*Sprachliche Kriterien*,” passim.

¹⁴ It should be noted that another German praxis, that of *Sprachstatistik*, was contemporary to Lutosławski but can properly be considered as a broader program that includes number theory; Lutosławski, “On Stylometry,” 284.

¹⁵ Lutosławski, *The Origin and Growth of Plato's Logic*; Lutosławski, “*Principes de stylométrie*,” passim.

¹⁶ Aghast that for twenty-eight years Campbell's work lay uncited, Lutosławski exclaimed, “Ces recherches de Campbell... sont... restées entièrement inconnues pendant vingt-huit ans et n'ont été citées par personne.” See Lutosławski, “*Principes de stylométrie*,” 62. Holmes was upbraided for the same slight in the review by Pawlowski and Pacewicz, “Wincenty Lutoslawski,” 423–24.

works and in a third abstract read before the Oxford Philological Society,¹⁷ laid out a paradigm that with little variation would come to characterize good experimental design practice for stylometry far in advance of the birth of modern inferential statistics. His program reads like a primer in simple design theory: samples of equal text size,^[f] large numbers of such samples,^[g] large numbers of markers,^[h] careful quantification of variation,^[i] and assignment of that variation to a cause.^{18[j]} Except for the telltale late-nineteenth-century predilection to discover “laws” whenever any regularities were observed, had Lutosławski been translated fifty years into the future and equipped with an updated statistical apparatus, he would have been virtually indistinguishable from the working stylometricians of the mid-twentieth century. For this and for his persistent efforts to position stylometry as an emerging science, Lutosławski should be granted pride of place as the inaugurator of the discipline.¹⁹

Now on to Mendenhall.²⁰ This scholar, as we have seen, began his stylometric excursions incited by de Morgan’s cogitations on word lengths.^[k] A note on de Morgan’s singular idea is appropriate here because it set the discipline on a sixty-year path in which *derivative* rather than *direct* characteristics of syntagms—their lengths—would occupy scholars rather than the syntagms themselves. This historic mistake can be excused, at the beginning of its course at least, because of its source, Augustus de Morgan. de Morgan was simply one of the nineteenth century’s most remarkable intellectual figures. By training and profession a mathematician, he maintained a lively correspondence across disciplines with a broad range of scholars and friends.

¹⁷ Lutosławski, “On Stylometry,” 284–86.

¹⁸ Lutosławski, “On Stylometry,” 284–85.

¹⁹ Lutosławski, however, credited Campbell with the original idea (Lutosławski, “*Principes de stylométrie*,” 62). For a similarly high estimate of the Polish scholar see Pawlowski and Pacewicz, “Wincenty Lutoslawski,” 440–41.

²⁰ Mendenhall, “Characteristic Curves of Composition,” *passim*.

In one such letter written to a Rev. Heald, dated August 18, 1851, de Morgan wrote the following:

... run your eye over any part of those of St. Paul's Epistles which begin with Παύλος—the Greek, I mean... Then do the same with the Epistle to the Hebrews, and try to balance in your own mind... whether the latter does not deal in longer words than the former.²¹

Though Mendenhall failed to recall correctly where de Morgan had actually said it²²—he nonetheless pursued de Morgan's suggestion. As a physicist moreover, Mendenhall chose to implement de Morgan's idea by producing a “spectra” (a histogram) of word lengths. This innovation allowed word-length histograms between texts or authors to be visually compared and overlaid upon one another,²³ a construct Mendenhall termed “curves of composition.” Two problems dogged the interpretation of his own curves, however. The first lay in his “mathematics of difference.” How much difference does it take, Mendenhall asked, to conclude that two curves are *truly* different? Statistics capable of answering this question simply did not exist in Mendenhall's day. Second, whether by the intuitive sense that other markers had more discriminatory power, or by the more linguistically obvious truism that words encode far more information by their lexematic variety than by their length,^[1] relatively few other researchers (save Williams in 1940 and 1956 and Brinegar in 1963) followed Mendenhall's lead in creating “characteristic curves.”²⁴

From this point onward, this review will use two terms to describe two distinct strata of work in stylistics. When speaking in just the classical, limited sense of using some kind of

²¹ de Morgan, *Memoir of Augustus de Morgan*, 215–16.

²² See Mendenhall, “Characteristic Curves of Composition,” 237.

²³ Mendenhall's initial 1877 work was followed by a later paper in 1901 that demonstrated that Mendenhall's own “curves” vary dramatically by language, and that Marlowe's characteristic curve was essentially indistinguishable from Shakespeare's (Mendenhall, “A Mechanical Solution of a Literary Problem,” 97–105).

²⁴ See Williams, “A Note on the Statistical Analysis of Sentence-Length,” 356–61; Williams, “A Note on an Early Statistical Study of Literary Style,” 248–56; Brinegar, “Mark Twain and the Quintus Curtius Snodgrass Letters,” 85–96.

mathematics to address the marquee historical focus of quantitative stylistics—authorship attribution—the classical term, stylometry, will be employed. When referring to the larger task, using computation to assist in assaying the whole complex of potential issues (sociocultural, dialectic, diachronic, etc.) that may drive stylistic differences between texts, Craig’s term, computational stylistics will be employed.²⁵ This usage is intentional because there is, in fact, a nested relationship between the two, with stylometry to be understood as a subset of computational stylistics.

3.3.2 Early Twentieth-Century Stylometric Descriptivism (1900–1935)²⁶

The early twentieth century set into motion four waves of research. As before, each will be treated in sequence.

3.3.2.1 Research Informed by Neither Linguistic nor Statistical Theory

Both late nineteenth-century and early twentieth-century stylometric research was handicapped by the misfortunes of arriving before Saussure’s linguistics (which provided a semiotic explanation for language with all the sociological frameworks that implies) and Fisher’s statistics, respectively. The complications that arose in the absence of these disciplines can be seen clearly by reviewing the work of two scholars of this period, Lucius Adelno Sherman, and Robert Édouard Moritz.

Sherman published two works in stylometry. The first of these explored sentence lengths following the path Mendenhall had forged the prior year. The second added to that work “present participles, past participles, and appositives.”²⁷ In both works, Sherman clearly uncovered a

²⁵ “Computational Stylistics aims to find patterns in language that are linked to the processes of writing and reading, and thus to ‘style’ in the wider sense, but are not demonstrable without computational methods.” Craig, “Stylistic Analysis and Authorship Studies,” 1.

²⁶ Delcourt accurately terms this quantitative but noninferential work as descriptive stylometry; Delcourt, “Stylometry,” 984.

²⁷ See especially Sherman, “Some Observations upon the Sentence-Length,” 119–30; Sherman, “Principles in the Development of Form in Literature,” 337–66. Sherman was aided in the latter work by his graduate student, G.A.

fascinating phenomenon: *something* was driving sentence lengths to become shorter over time.^{28[m]} Unfortunately, Sherman did not, as one might imagine, collect a range of possible idiolectic and sociolectic causes for this stylochronometric finding. Rather, he centered his inquiry on whether authors “as eminent and practiced as these, are subject to a rigid rhythmic law from which...they may not escape.”²⁹ *For Sherman this was surely authorial variation...and the de facto cause was a rigid law.*^[n] From our twenty-first-century perspective given the same data, we would surmise that such phenomena are explicable by any number of causes. Yet, this conclusion misunderstands the times in which Sherman lived. This was not an era amenable to considering whether linguistic directionality over time was guided by colligative social forces from the *context of culture* in which we learn and adapt our own speech patterns to the ever-changing speech-world that surrounds us. Accordingly, rather than interpreting these broad sociocultural changes socioculturally, or perhaps in terms that Grice would later suggest,³⁰ Sherman interpreted them as demanded by his priors. Sherman’s hermeneutic was solidly situated within a perspective that considered complex phenomena through the lens of deterministic mathematics, a view in aggressive ascent when Sherman wrote. This ascent, in fact, would only reach its high water mark thirty years after Sherman when Hilbert attempted to find a mathematical theory to explain all of mathematics.³¹ Importantly, for the thesis of this work,

Gerwig, who would later publish his own work on sentences (Gerwig, “On the Decrease of Predication and of Sentence Weight,” 17–44).

²⁸ Subsequent research has demonstrated that there is nothing quite as irresistible as measuring what is currently in vogue, in this case, sentence lengths. Mosteller and Wallace would later discover sentence lengths to be all but useless (Mosteller and Wallace, “Inference in an Authorship Problem,” 276).

²⁹ Sherman, “Principles in the Development of Form in Literature,” 348.

³⁰ Grice articulated the principle of cooperation; we act and speak *so as to be understood*. See especially Grice, “Logic and Conversation,” 39.

³¹ This was Hilbert’s *Entscheidungsproblem* that enquired whether an algorithm could be constructed to decide if the provability of a given statement existed if the predication of that statement was the correct and complete rules of logic alone. See especially Hilbert’s rationale and hope for such a program; Hilbert, “Die Grundlegung der elementaren Zahlenlehre,” 485–89.

what mathematics delivered instead was Gödel's incompleteness theorem: proof that given any finite number of axioms, true statements exist about mathematics that cannot be proved from those same axioms. This meant only one thing—mathematics was likely not complete in itself. The implication was stunning; *openness to perturbations outside the system was demanded by the nature of the system itself*. These are thoughts disquieting, however, to the nineteenth- and early twentieth-century materialistic mind.³² Despite the distaste that mechanistic materialists would later express for it, twenty years after Sherman, Gödel's discoveries and other discoveries like it would rewrite the narrative of mathematical self-understanding that so influenced the stylometric cogitations of Sherman and Moritz.³³ Finally, by the end of the 1930s these discoveries would sweep aside the presumed mathematical foundations of deterministic materialism.³⁴ Yet Sherman and Sherman's world was hardly ready for such ideas. Plurality of causation, social or otherwise, simply undermined certainties and violated the principle of elegance that late twentieth-century mathematical materialism had so successfully delivered over to its willing subjects. Thus, Sherman assumed, *de facto*, that "rigid laws" operated—and at the level of authorship. This orientation, discovering an apparent significant difference and assigning it to authorial causation without first eliminating potential alternative causes,^[o] would come to characterize much of stylometric praxis for the next eighty years.³⁵ Even granting Sherman

³² The hermeneutical implications of this are also profound. If understanding demands openness to extrasystemic intrusions, this has implications not merely in the plurality of perspectives (which may or may not follow from the mathematics) but also in the possibility of the incompleteness of physical systems without consideration of the metaphysical causes. More importantly it may challenge the division itself as vacuous.

³³ The nail on the coffin of determinacy came, finally, not through Hilbert's failed program, or Gödel's incompleteness theorem, but courtesy of Heisenberg's uncertainty principle. This triad—theoretical incompleteness, physical indeterminacy, and statistical uncertainty—suggested or demanded that understanding a system demands *externality*, the possibility of intrusion, and lastly that that intrusion itself (e.g. the observer in the Copenhagen interpretation of quantum mechanics) can change the state of the system *nondeterministically*.

³⁴ See the discussion on the rise, perigee and fall of early-twentieth-century positivism in Section 3.8.

³⁵ Recall that no science of inferential statistics existed as of Sherman's writing.

some relief because of the pre-linguistic times in which he lived, we might suggest that his own data could—and should—have led him to posit alternative sociological causes.³⁶

Our second figure, Robert Édouard Moritz, is seldom mentioned in reviews of computational stylistics. Yet, his contribution is critical because he presaged two movements which would soon emerge in stylometric practice. First, in his 1903 *University Studies* article, he exemplified, precociously, the full rhetorical flower of high positivism,³⁷ replete with legitimating his findings by reference to self-formulated “principles” and “laws.”³⁸ Second, wholly unencumbered by an intuition of the nuanced complexity of language, Moritz successfully evangelized for an unmitigated form of linguistic synecdoche, that is, that a single linguistic measure (such as sentence-lengths) could with “no uncertainty, no indefiniteness, and no ambiguity”³⁹ serve as a stolid test for authorship. This confidence was predicated on his unwavering conclusion that “it is not possible for an author to escape his stylistic tendencies.”⁴⁰ By these two summary tendencies, Moritz went well beyond Sherman. Here is how he did it. After quoting Sherman extensively, he canonized Sherman’s far more nuanced arguments as “Sherman’s principle”⁴¹ ebulliently proposing that sentence rhythm was the “hidden mark, the cipher...with which each author unconsciously endows the progeny of his pen.”⁴² Next, claiming that Sherman’s initial work was more than sufficient to show that “it is not possible for a writer

³⁶ Specifically, Sherman’s own data demonstrated that the shift to shorter sentences over time was disrupted by “dialog passages and consequent reduced averages” (p. 16) and shifts to an “oral sentence sense” from the older “literary sentence form” (p. 23). Both are naturally evocative of generic causes. Sherman, “Principles in the Development of Form in Literature,” 12–23.

³⁷ Cashdollar’s thesis should be engaged here—that by the late nineteenth century positivism was all but dead (Cashdollar, *The Transformation of Theology*, 163). What was true in the halls of philosophy, however, was decidedly not true in the halls of science. Even as Moritz wrote, the mathematical enterprise via Whitehead, Russell, and Hilbert were hard at work fitting positivism for its definitive reinstatement in the Vienna and Berlin circles. See section 3.8.1 for the stultifying effect this had upon computational stylistics.

³⁸ Moritz, “On the Variation of Certain Sentence-Constants,” 233.

³⁹ Moritz, “On the Variation of Certain Sentence-Constants,” 234.

⁴⁰ Moritz, “On the Variation of Certain Sentence-Constants,” 234.

⁴¹ Moritz, “On the Variation of Certain Sentence-Constants,” 233.

⁴² Moritz, “On the Variation of Certain Sentence-Constants,” 229–30.

to escape from his stylistic tendencies,” he was then free to confidently assert the inviolable connection between sentence lengths and authorship. Intriguingly, however, Moritz’s own work even granted the opposite, concluding just as securely that “sentence-lengths...are quite as much dependent upon the nature of the composition...as upon the author’s sentence instinct.”⁴³ The careful reader, then, is forced to ask, “Which is it?” Moritz sensed his predicament but still failed to fully extricate himself from his first overstatement.⁴⁴ Subsequent research now makes it far less difficult to disambiguate Moritz’s findings. What Moritz indeed discovered was that *some* authors are distinguishable from others when writing in the same genre and that almost all authors display different sentence lengths when writing in different genres. Despite the collisions of his overstatements,⁴⁵ Moritz’s own data moved the discipline forward, but in a way he did not expect, by demonstrating that even an indicator as coarse as sentence lengths could reveal patterning not only by author, but even more extensively by genre. This genre-centric notion was only dismissed by Moritz because multiple causality fit cleanly inside neither Sherman’s nor his own brand of linguistic determinism.

3.3.2.2 Research Informed by Quantitative Theory but not Linguistic Theory

Research informed by some variety of statistical or quantitative theory but absent a formal linguistic apparatus would include the classic articles on letter sequences by A. A. Markov. Markov’s contribution lay in applying his eponymous “Markov chains” to texts. This was achieved twice: first in 1913⁴⁶ and then in his posthumously published *Calculus of Probabilities* in 1924.⁴⁷ Two of his findings are relevant for our purposes. First, implicit in

⁴³ Moritz, “On the Variation of Certain Sentence-Constants,” 239.

⁴⁴ See his footnote (Moritz, “On the Variation of Certain Sentence-Constants,” 243).

⁴⁵ Craig speaks to this, “A well-founded computational stylistics works with tendencies rather than rules.” Craig, “Stylistic Analysis and Authorship Studies,” 4.

⁴⁶ Markov, “Пример статистического,” *passim*.

⁴⁷ Markov, *Ischislenie veroiatnostei*. Almost seventy years later Petruszewycz resurrected Markov chains in two comparison articles (Petruszewycz, *Les chaînes de Markov*; Petruszewycz, “Chaînes de Markov,” 85–95). Her work

Markov's mathematics is the formalization of a system that transitions from one state to another. These transitions are, intriguingly, probabilistic and *not* deterministic.⁴⁸ Moreover, the system changes randomly, though within a probability distribution where transitions from one state to another are not required to be equiprobable.⁴⁹ Accordingly, Markov *mathematics was open to multiple causation* and not surprisingly, became one of the first varieties of mathematics used to model sociolectic rather than idiolectic outcomes—the very thing that Sherman and Moritz delegitimated.⁵⁰ Second, while later research would confirm that Markov chains could classify either sociolectically⁵¹ or idiolectically⁵² grouped texts to a high degree of correct classification, Markov models do not seem to substantially outperform other supervised techniques.⁵³ As such they will not play a large role in the articulation of our experimental design in Chapter Four.

3.3.2.3 NT Stylometric Descriptivism and the Utility of *Hapax Legomena* (1870–1922)

Four scholars, Gardner (1887),⁵⁴ Parry (1920),⁵⁵ Penick (1921)⁵⁶ and Harrison (1921)⁵⁷ made contributions to NT stylometry in the period between 1870 and 1923. Section 1.4 already introduced in some detail the contribution of Harrison and Parry. For the purposes of this review we will therefore limit ourselves only to the most salient issue raised by these four scholars—and to no New Testament scholar's surprise, it is the issue of *hapax legomena*. It best serves this

demonstrated not only the utility of Markov chains, it arguably elevated Markov chains alongside Information Theory as two of the most potent and underutilized techniques in quantitative stylistics.

⁴⁸ See the discussion in 3.3.2.1.

⁴⁹ These and other properties of Markov processes are discussed in the first three chapters of their work (Meyn and Tweedie, *Markov Chains*, 1–79).

⁵⁰ Petruszewycz, *Les chaînes de Markov*, 40–74.

⁵¹ Tsimboukakis and Tambouratzis, “Word-Map Systems,” *passim*.

⁵² Khmelev and Tweedie, “Using Markov Chains,” *passim*; Rand, “Markov Models of Literary Style,” *passim*; Putniņš et al., “Advanced Text Authorship Detection Methods,” *passim*.

⁵³ Brainerd and Chang, “Number of Occurrences in Two-State Markov Chains,” 231; Tsimboukakis and Tambouratzis, “Word-Map Systems,” 67; Putniņš et al., “Advanced Text Authorship Detection Methods,” J-12.

⁵⁴ Gardiner, “The Language of the Epistle to the Hebrews,” *passim*.

⁵⁵ Parry, *The Pastoral Epistles with Introduction*, *passim*.

⁵⁶ Penick, “Paul's Epistles Compared with One Another,” *passim*.

⁵⁷ Harrison, *The Problem of the Pastoral Epistles*.

topic to treat it diachronically and then to compare and contrast Parry and Harrison in particular with subsequent research.

First, recall from Section 1.4 that Parry argued that the novel vocabulary expressed in the Pastorals was due to *subject matter*.⁵⁸ Harrison, in contrast, argued that its signature in the NT left little room to conclude anything other than a second-century source: a *pseudepigraphal author* who used genuine fragments from the apostle.⁵⁹ Given these distinct stances, how have these disparate conclusions about *hapax legomena* stood the test of time? Four things can be said about Harrison's *hapax legomena* in terms of subsequent research:

1. *Hapax legomena* have fallen from favor in authorship studies.⁶⁰
2. *Hapax legomena* and low-frequency words in general are now seen to be far more associated with *genre or other components of sociolect* rather than authorship.⁶¹
3. High-frequency words are now definitively accepted as better markers of authorship than low-frequency words.⁶² The two most successful approaches in the subsequent history of research in authorship attribution have not been based on low-frequency measures but *high-frequency* measures.

⁵⁸ Parry, *The Pastoral Epistles with Introduction*, cxxvi.

⁵⁹ Harrison, *The Problem of the Pastoral Epistles*, 95–102. It should be clarified that Harrison's method of counting *hapax legomena* was deeply flawed per the separate assessments of Grayston, Michaelis, Guthrie and O'Donnell. Michaelis, "Pastoralbriefe und Wortstatistik," 69–76; Grayston and Herdan, "Authorship of the Pastorals," 4–5; Guthrie, *New Testament Introduction*, 67; O'Donnell, "Linguistic Fingerprints," 209.

⁶⁰ The issue boils down to a misconception. *Hapax legomena* are not necessarily idiosyncratic – they arise naturally in shifts of register in our day-to-day life. Second, the inappropriateness of *hapax legomena* as a clear marker for authorship is now granted on almost all sides; hence Greenspahn, "The Number and Distribution of Hapax Legomena," 18; Mosteller, "The Writing Styles of the Authors of 'the Federalist' Papers," 140. Mosteller terms frequencies of less than five as "nearly useless for discrimination"; Mosteller and Wallace, "Inference in an Authorship Problem," 280. The question becomes quite different, however, when (a) low-frequency words constitute the largest percent of the total number of words and (b) the texts are not from a mixed-genre corpus. For this condition, see Labbé, "Authorship Attribution by Intertextual Distance," 48.

⁶¹ We have already seen Greenspahn comment that "Poetry uses more *hapax legomena* than prose." Greenspahn, "The Number and Distribution of Hapax Legomena," 18–19. Moreover, very often infrequent words are register-specific, as noted in Campbell, *Sophistes and Politicus*, xxxiii. Regarding the mathematical relation between *hapax legomena* and dating see Honoré, "Some Simple Measures of Richness," 172–77. Quite importantly, despite the depreciated status of *hapax* in authorship studies, it would be just as mistaken to conclude that they are without value in *any* classification studies. Stamatatos demonstrates that rare words discriminate text-types second best out of twenty-two measures (Stamatatos et al., "Automatic Text Categorization," 492). This, again, is concordant with register theory which almost tautologously states that special words are found in special registers.

⁶² Per Smith, "*hapaxes*....are of dubious value as indicators of authorship." Smith and Rickards, "The Authorship of the American Declaration of Independence," 21. See also Binongo and Smith, "Statistical Approaches to Philippine Literature," 522.

- a. The high-frequency “function words” of the “Burrows’ method” (late 1980s to present).⁶³
- b. The high-frequency largely lexical “feature-vectors” used by the current generation of supervised authorship attribution methods (mid-1990s to the present).⁶⁴

Clearly, all four findings above, if found to be true of NT texts as well, would decidedly tilt the computational stylistics argument far in Parry’s favor. *It is important to note, however, that these four general findings have yet to be confirmed in anything remotely approaching a comprehensive experimental design in the GNT.*⁶⁵ Mitigating this fact, however, is that one aspect of register theory seems unavoidably ubiquitous—we all use specialized vocabulary in specialized settings.⁶⁶ Novel vocabulary is the natural byproduct of a distinct *context of situation* because it calls forth a different register. This is no arid theoretical construct, since it has been experimentally demonstrated, per point two above.⁶⁷ It is quite difficult to see how the NT would be an exception to such a sociolinguistically ubiquitous finding. These factors suggest that both

⁶³ In discussing the now accepted greater utility of high-frequency words, Burrows explains that high-frequency function words work so well as discriminators because they “provide the main strands of the whole fabric.” Burrows, “A Second Opinion,” 377. Dozens of studies have adopted the “Burrows Method,” the use of high-frequency function words, as the best overall discriminators of authorship (e.g. Forsyth et al., “Investigating the Authenticity of the Consolatio,” 379–82; Hoover, “Multivariate Analysis and the Study of Style Variation,” 341–42).

⁶⁴ Argamon et al., “Stylistic Text Classification,” 808; Smith and Rickards, “The Authorship of the American Declaration of Independence,” 21; Luyckx, *Scalability Issues*, 28; Kestemont et al., “Robust Rhymes,” 60–62.

⁶⁵ A more full-throated approach to an experimental design would first explore a range of low-frequency words (one, two, through N occurrences) in GNT texts, then do the same for GNT texts normed for size (per Zipf’s law). Next, it would perform both these exercises by *a priori* GNT text groupings such as the authorship and genre groupings described in Chapter Five of this work. (This would include any other hypothesized sociolectic or idiolectic text grouping as well.) Last, frequencies normed to the distribution of all GNT texts and hypothesized text groupings should also be explored, that is, by bottom deciles, centiles, etc.

⁶⁶ While glossematic, systemic and stratificational linguistics all view register differently (Matthiessen, “Register in the Round,” 224) the common ground between these theories seems to be that specialized vocabulary is employed in specialized settings. Perhaps the best general moniker for this is Firth’s notion of “restricted languages.” See de Beaugrande, “Register...in Search of a Theory,” 8.

⁶⁷ The excellent discrimination provided by special (low-frequency) vocabulary in modern Greek, for instance, certainly seems to lean in that direction (Stamatatos et al., “Automatic Text Categorization,” 492).

low- and high-frequency linguistic measures (not just *hapax* alone) should be tested in our experimental design.^{68[p]}

3.4 Milestone One: Modern Linguistics Emerges (1890s–1920)

Historical and general introductions to linguistics abound, such as the eminently readable historical work by Samson,⁶⁹ the earlier and more technical general works by Lyons, Davis, and Lehmann,⁷⁰ as well as the ever-evolving Cambridge Textbooks in Linguistics series.⁷¹ Each of these in different ways has signaled that a seismic shift in language theory occurred in the early decades of the twentieth century. This shift was due in large part to one well-known figure, Ferdinand de Saussure, and one far lesser-known one, Franz Boas.⁷² The first was a structuralist and theoretician in France, the second an anthropologist and, eventually, a descriptive linguist in America. It would take most of a century and several convulsive periods in linguistics (the last one called, without exaggeration, the linguistic wars) to work through the insights raised by these two men. In the process, however, both Saussure's structuralism and Boas' descriptivism would perish, albeit with certain central ideas of each inherited by the dominant forms of linguistics today: Transformational Generative Grammar in North America, and the various functional

⁶⁸ Regarding *The Problem of the Pastorals*, scarcely has any work on historical style been so winsomely written, logically argued, fully conversant with alternative causes, and yet so thorough in dismissing the weight of those same causes. Harrison's brilliance in his work, and there was much of that on display, lay not in the novelty of a quantitative approach to the topic, nor the novelty of his late dating of the texts, nor even the idea for which he is best known—the utility of *hapax legomena* as a fingerprint of style. His brilliance lay in interweaving all three with other grammatical forms besides into a narrative of indisputable discursive weight. As our review demonstrated above, however, his methodological assumptions have not weathered well at all.

⁶⁹ Sampson, *Schools of Linguistics*.

⁷⁰ Lyons, *Introduction to Theoretical Linguistics*; Davis, *Modern Theories of Language*; Lehmann, *Descriptive Linguistics*.

⁷¹ See especially the first two chapters in Jeffries' work on stylistics (Jeffries and McIntyre, *Stylistics*, 1–65).

⁷² Here we essentially ignore the trajectory traced by the philologists and comparative philologists whose ideas on language theory were felt deep into the twentieth century. See section 2.5.4 for a quick survey of the philologists von Schlegel, Humbolt, Grimm, Bopp, Delbrück, and Ferrar. We ignore, too, the work of Sapir and Whorf. For the contribution of these and their eventual pairing see Robins, *A Short History of Linguistics*, 236–37; Halliday, *Halliday: System and Function in Language*, ix–x; Butler, *Systemic Linguistics: Theory and Applications*, 225; Dixon, *What is Language*, 97–100. For Sapir's anticipation of the emic-etic distinction so central to Pike's tagmemics, see Pike, *Language in Relation to a Unified Theory*, 39, 53–55.

schools in Europe.⁷³ It does not serve our immediate purpose to provide a deeper introduction to linguistics than this, save making one last point. In a persistent way, and one that seems frequent in the evolution of disciplines, the dichotomy between the “top-down” theorizing of Saussure, and the “bottom-up” descriptivism of Boas has lived on in the TGG of Chomsky and the Systemic Functional Grammar of the London school respectively.⁷⁴ Rather than debate the merits of each, our concern is to find that school of linguistics that will most adequately deliver language measures to help decode the information encoded in language. We have already covered in Section 2.5 the admittedly contentious discussion of what language school that may be. The central notion that compels us is that if quantification is to cooperate with language theory—as it ought—it *must* operate upon actual *language-in-use*, not upon disembodied intuitive theories. Against the intuitive notions of Transformational grammar, Sampson’s rationale for such a commonsense approach is both terse and refreshing;

If intuitions shared by the leaders of the [linguistic] discipline could get the facts of language as wrong as this, it is imperative to find some other way of engaging with the concrete empirical realities of language... Happily, easy access to computers...and corpora had arrived just in time to solve this problem.⁷⁵

As a consequence of this necessary move to the empirical study of actual language-in-use, we arrive at following initial set of design implications.

3.4.1 Design Implications

1. *The functional schools in general and Systemic Functional Linguistics (SFL) in particular provide a far richer set of language measures than traditional grammar.*⁷⁶

⁷³ Regarding the functional schools in their several varieties and the Transformational Generative Grammar of Chomsky, only the former are concerned fundamentally with what Sampson quite rightly terms, “empirical linguistics.” See especially Sampson, *Empirical Linguistics*, 1–12.

⁷⁴ For a summary schema of the relationship between theory and praxis in SFL see Teich, *Systemic Functional Grammar in Natural Language Generation*, 3–5.

⁷⁵ Sampson, *Empirical Linguistics*, 22.

⁷⁶ See section 2.5.

- a. Implication: Use SFL-derived linguistic measures as a starting point for the design.^[a]
2. *The langue/parole distinction as a construct in computational stylistics is essentially vacuous. Only parole, language-in-use, has relevance to computational stylistics.*
 - a. Implication: Use principled collections of texts as they are actually spoken and written. Our eventual scope must be all of the Koine.^[b] For pragmatic reasons, however, we must limit this current work to the mixed-genre texts of the NT.^[c]
 3. *Substantial advances in semantic domain theory have been achieved.*
 - a. Implication: Explore strata other than the “wordings” of lexicogrammar. Include the “meanings” in the semantic strata in language.^[d] Those domains should eventually be decided by either proximity methods of semantics or the calculation of semantic domain models rather than intuitively.⁷⁷ For now the domains of Louw & Nida will suffice as an initial point of departure.⁷⁸
 4. *Beyond general descriptive stylistic studies in the GNT,⁷⁹ the linguistic notions of the context of culture and context of situation^{80[e]} must be more thoroughly explored.⁸¹*

Implication: Further explore sociolect by (1) defining the domain of sociolect in terms of context of culture (e.g. diatypic, dialectic, economic/class etc.) and context of situation (e.g. genre, date, gender, register, audience, etc.) and (2) extracting structural components

⁷⁷ Gliozzo and Strapparava, *Semantic Domains in Computational Linguistics*, 13–48.

⁷⁸ Louw and Nida, eds., *Greek-English Lexicon of the New Testament, Volume 2*.

⁷⁹ Such as the widely separated comparative stylistic works of Simcox and Spenser; Simcox, *The Writers of the New Testament: Their Style and Characteristics*, 90–154; Spenser, *Paul's Literary Style*.

⁸⁰ For these contexts see Halliday, *Halliday: System and Function in Language*, viii–xv; Halliday and Matthiessen, *An Introduction to Functional Grammar*, 26–29; Butler, *Systemic Linguistics: Theory and Applications*, 4–6.

⁸¹ For quantitative work in the GNT sensitive to modern linguistics see Black, *Sentence Conjunction in the Gospel of Matthew*, 95–101; Kwong, *The Word Order of the Gospel of Luke*, 46–128; Martín-Asensio, *Transitivity-Based Foreground in the Acts of the Apostles*, 94–110.

of variation^[f] and probing for any existing mathematical relation⁸² between that variation and the elements of the domains just defined. These will require, among other things, the various flavors of extractive multivariate mathematics that were detailed in section 2.5.

3.5 Milestone Two: The Birth of Statistics and Quantitative Linguistics (1922–1935)

Mathematical description and mathematical inference respectively are not merely fundamental tools of the modern quantitative scholar, they are constitutive of what the quantitative scholar does and why she or he does it. Without this pair, the modern quantitative disciplines, at least as we know them, would not exist. In the period we now review (1922 – 1935), two central developments made quantitative *linguistic* scholarship, including its computational stylistic branch, possible. The first was the *mathematics of inference* and hypothesis testing itself: modern classical statistics. The second was the inauguration of the *mathematics of linguistic description*.

3.5.1 R.A. Fisher and the Birth of Classical (Inferential) Statistics

Few who regularly use inferential statistics realize how young a science it really is. A demonstration of this can be had even within the boundaries of our own review as Lutosławski simply proposed (guessed) that a sufficient lexical difference would be “10% between two works”—but immediately equivocated by saying, “even this being in some cases insufficient.”⁸³ Lutosławski’s uncertainty was no reflection on him. Quite simply, in the late nineteenth century when questions arose regarding *how different* measures must be to be considered “truly”

⁸² To be specific, these relations are univariate, bivariate, multivariate or causal. See Hartwig and Dearing, *Exploratory Data Analysis*, 69–75.

⁸³ Lutosławski, “On Stylometry,” 285.

different,⁸⁴ no one really knew.⁸⁵ No one, that is, until one man derived—essentially *de novo*—the mathematics of statistical inference. That man’s name was R.A. Fisher.

Ronald Aylmer Fisher, in company with few others in this review, was a true intellectual giant, a polymath who in no particular order made important contributions to biology, genetics, mathematics, statistics, and, far less fortunately, to eugenics. It is, however, his central role in statistical theory that concerns us here. Of that, his contribution is clear. Hald termed Fisher without the slightest hint of hyperbole, “a genius who almost single-handedly created the foundations for modern statistical science.”⁸⁶ The achievement to which Hald refers began with Fisher’s publication of a simple, lucid and brief article, “On the Mathematical Foundations of Theoretical Statistics” (1922). This was followed by book-length works, *Statistical Methods for Research Workers* (1925) and *Design of Experiments* (1935).⁸⁷ The latter books would not only cement his reputation as the father of modern classical “Fisherian” statistics, but would eventually result in his knighthood.⁸⁸

For purposes of this review Fisher proposed four things that collectively defined the modern discipline and practice of classical statistical inference:

1. *Statistical differences must be measured against some standard of measure.* For these he proposed various distributions based on the types of data.⁸⁹
2. *It is vacuous to claim statistical difference using measures of central tendency alone.* By this Fisher announced that in addition to a difference in central tendency (e.g. the mean), the variance or “spread” of that distribution must also be assayed.⁹⁰

⁸⁴ More technically, until 1922 no formal theory existed to tell researchers if a single number, or set of numbers, was distant enough from some putative population to be statistically significant.

⁸⁵ Accordingly, it is accurate to conclude that all quantitative inference prior to Fisher was merely comparative and descriptive.

⁸⁶ Hald, *A History of Mathematical Statistics*, 738.

⁸⁷ Fisher, “On the Mathematical Foundations of Theoretical Statistics,” *passim*; Fisher, *Statistical Methods for Research Workers*; Fisher, *The Design of Experiments*.

⁸⁸ *Pace* Bayesian inference which Fisher opposed (Fisher, “On the Mathematical Foundations of Theoretical Statistics,” 310–11).

⁸⁹ So-called problems of distribution (Fisher, “On the Mathematical Foundations of Theoretical Statistics,” 311–13).

⁹⁰ So-called problems of estimation (Fisher, “On the Mathematical Foundations of Theoretical Statistics,” 316–17).

3. *It is vacuous to claim statistical difference without first defining conclusory error.* In this Fisher required that one must first define *the extent to which one is willing to be wrong* before any assertion of statistical difference. (This is the same as defining the likelihood of Type I error.)⁹¹
4. *Statistical difference is confirmed by rejecting the hypothesis of equivalence.* Statistical inference should operate by means of establishing an hypothesis of equivalence. Fisher's formal tests of inference, then, test whether such "null-hypotheses" should be accepted or rejected.⁹²

3.5.2 Zipf and the Birth of Quantitative Linguistics

The most notable early practitioner sensitive to both linguistic and quantitative theory was George Kingsley Zipf. As both a linguist and head of the German department at Harvard, Zipf pursued stylometric concerns as a tool to solve problems and quandaries in linguistics. As a scholar conversant with mathematics (he held the post of university lecturer) he could—and did—think mathematically about linguistics problems, eventually publishing his findings in two relatively brief book-length works.⁹³ Overlooked by many, Zipf's contributions are precocious in terms of experimental design and consequently deserve far more attention than we can give them here.⁹⁴ We have time to discuss only four of those contributions. First, *experimentally*, Zipf began where experimental design must, with the forest—the whole distribution of any given language category—rather than selected trees or “markers.”^{95[a]} Second, *linguistically*, he demonstrated that different language features display different distributions.⁹⁶ This, in turn,

⁹¹ In Fisher's paper this was achieved by deriving the standard error (Fisher, “On the Mathematical Foundations of Theoretical Statistics,” 323–30) and then exploring the errors of the moments of the normal distribution (Fisher, “On the Mathematical Foundations of Theoretical Statistics,” 338–55).

⁹² See Fisher, *Statistical Methods for Research Workers*, 80–159.

⁹³ Zipf, *Selected Studies of the Principle of Relative Frequency*; Zipf, *The Psycho-Biology of Language*.

⁹⁴ See for instance, Zipf's articulation of linguistic markedness without using the term (Zipf, *Selected Studies of the Principle of Relative Frequency*, 1).

⁹⁵ Hence he looked at all word distributions from the least to the most frequent (Zipf, *Selected Studies of the Principle of Relative Frequency*, 22–27).

⁹⁶ Such as the phonemic variance across different languages (Zipf, *Selected Studies of the Principle of Relative Frequency*, 2).

provided empirical support for the emergent linguistic notion that language was paradigmatically and syntagmatically multifaceted, nuanced, and complex. Third, *quantitatively* his mathematics led him to the development of *Zipf's law*, a relation that describes the shape not only of word distributions but the distributions of many physical and social systems.⁹⁷ Fourth, to my knowledge, he was the first quantitative researcher to characterize and discuss at length the strata of semantics via “semantic shift/drift”—the semantic plasticity of lexis over time.⁹⁸

3.5.3 Design Implications

Four implications follow from these two quantitative milestones.

1. Classical Fisherian statistical inference should be used both for parametric (discontinuous) and nonparametric (continuous) data.^[b]
2. Other subsequent methods of statistical inference should be tested in parallel alongside classical Fisherian statistics.⁹⁹ The rationale that demands this is our stated commitment to a fully coherentist approach, and because Fisherian statistics is inveterate in the use of putative distributions and hence essentially foundationalist in its priors. That is, it assumes that the sample is drawn from the putative distributions of Fisherian statistics. Accordingly, three other methods of inference should be employed:
 - a. Bayesian inference: a distribution-free approach to inference^{100[c]}
 - b. Bootstrapping: a second distribution-free approach to inference that develops its distribution up from the data of the sample itself.^{101[d]}
 - c. Mixed methods approaches to inference and validation should also be employed for the same reason.^{102[e]}
3. Zipf's insight that quantitative description begins with the whole distribution should be followed.
4. Zipf's work was the first to give credence to the notion that whole-heeled linguistic description must not only embrace *grammar* and *pragmatics* but also *semantics*. This should be appropriated into our experimental design as well.^[f]

⁹⁷ That is, $F_n \sim \frac{1}{n^p}$ where F_n is the frequency of a word of rank n , and the exponent p is close to 1.0.

⁹⁸ Zipf, *Selected Studies of the Principle of Relative Frequency*, 8–15.

⁹⁹ For the initial example of the this kind of work, see Mosteller and Wallace, “Inference in an Authorship Problem,” 282–302.

¹⁰⁰ See Iversen, *Bayesian Statistical Inference*, 7–17.

¹⁰¹ Mooney and Duval, *Bootstrapping: A Nonparametric Approach*, 1–14.

¹⁰² See especially, Tashakkori and Teddlie, *Handbook of Mixed Methods*, 285–313.

3.6 Period Three: Harps in Babylon: Description without Inference (1922–1960s)

3.6.1 The Three Research Trends During the Period

The period of twentieth-century stylometric descriptivism (Section 3.3) demonstrated that competing schools animated by different methodological approaches operated side by side in computational stylistics, as indeed they still do in many disciplines today. So it was in the period after the quantitative achievements of Fisher and Zipf. One movement, for instance, appropriated Fisher's insights immediately as laid out in his *Manual for Statistical Workers*. The other summarily ignored it. This latter choice inaugurated a period that might somewhat pejoratively but accurately be described as a period of description without a difference. Most of the work during this time used only a small set of *ad hoc* markers. Works that chose both small numbers of *ad hoc* markers¹⁰³ absent even the most rudimentary tests of statistical significance included the work of Jefferson on Isaiah (1949),¹⁰⁴ P.N. Harrison (1955) on the Pastoral Epistles (a reassertion of his 1921 dissertation),¹⁰⁵ and Sanders' (1966) attempted renovation of Holtzmann's original opinion on Colossians.¹⁰⁶ Outside of biblical studies the same dual tendencies can be found in Bennet's (1957) analysis of style in two disparate plays of Shakespeare,¹⁰⁷ Hartwell's (1970) otherwise helpful word counts on Hardy,¹⁰⁸ Frautschi's (1967)

¹⁰³ The quantitative treatments embraced only simple moments of the respective distributions, lists of frequencies or means, and the like.

¹⁰⁴ Jefferson following Köhler used three *ad hoc* markers—the use of the article, the sign of the accusative (אָ), and the omission of the relative pronoun אשר—to contribute to the authorship debate regarding Isaiah 65 and 66. While Jefferson concludes that these markers are generally unfavorable to assigning Isaiah 66 to either Deutero or Trito-Isaiah, the lack of significant sample, the *ad hoc* selection of markers, and the complete absence of an inferential architecture, all mark Jefferson's work as a collision of design flaws wholly inadequate to support any claimed conclusions (Jefferson, "Notes on the Authorship of Isaiah 65 and 66," 227).

¹⁰⁵ Harrison, "Important Hypotheses Reconsidered," *passim*.

¹⁰⁶ For his decidedly non-inferential quantitative argument see Sanders, "Literary Dependence in Colossians," 28–31.

¹⁰⁷ Bennett at least used Yule's *K*, recognizing its nuanced utility but exercised upon it no tests of inference (Bennett, "The Statistical Measurement of a Stylistic Trait," 33–50).

¹⁰⁸ Hartwell, "A Quantitative Approach to Thomas Hardy's Prose Style," 153.

authorship work on Diderot's *Encyclopédie*,¹⁰⁹ Maloney et al.'s (1977) analyses on Hippocrates,¹¹⁰ and Davison's (1984) comparison between Paul and Luke.¹¹¹

Other research, in contrast, assembled large sets of linguistic markers but, as before, analyzed then completely apart from what was by then *de rigueur* methods of classical Fisherian inference. Examples include the work by Baird (1976) on content analysis¹¹² and Weingartner (1971) on the putative Old French Lais of Marie de France.¹¹³

Other practitioners whose work displays fundamental historical, literary, linguistic and/or mathematical errors obvious even in their own time will be omitted from our experimental design review.¹¹⁴

3.6.2 Looking for the Linguistic Silver Bullet

Researchers during the early part of the twentieth century began a more aggressive attempt to press beyond word lengths, sentence lengths, and *hapax legomena* in their search for more discriminatory stylometric measures. The principled attempt to secure better measures, was, to be sure, driven in part by fundamental research. Yet, it also filled a deep stylometric need. Given that sentence lengths, a scalar measure, had demonstrated at least *some* discriminator power, it seemed reasonable to hope that a more refined key to authorial discrimination lay in other scalars. Might, then, the *parameters* of an author's own language distribution contain some hidden key, the hoped-for cipher of Moritz? This reasoning contributed

¹⁰⁹ Frautschi, "A Project for Author Discrimination," 14–17.

¹¹⁰ Maloney et al., "*La longueur des mots*," 95–122.

¹¹¹ Davison, "Paul v. Luke," 1–4.

¹¹² Baird, "Content Analysis and the Computer," *passim*. Baird encourages, "The testing of content validity... for mathematical probability," but then, in an odd omission, does none himself (Baird, "Content Analysis and the Computer," 273).

¹¹³ Weingartner, "Stylistic Analysis of an Anonymous Work," 1–9.

¹¹⁴ One notable exception should be made, the work of Marriott, who thought nothing of plotting a measure of central tendency against its own variance orthogonally and then secured his own high opinion of the error by labeling each plot with the phrase "Copyright reserved" (Marriott, "The Authorship of the *Historia Augusta*," 67–72).

to a thirty-year project in which dozens of researchers either developed measures of linguistic diversity or utilized those produced by the early fundamental forays in quantitative linguistics.¹¹⁵ These measures included, in no particular order, metrics such as Yule's K,¹¹⁶ the type/token ratio,¹¹⁷ Simpson's index,¹¹⁸ the Sichel distribution,¹¹⁹ Greenberg's index of linguistic diversity,¹²⁰ entropy,¹²¹ Guiraud's index of vocabulary richness¹²² as well as various rank distribution measures,¹²³ punctuation,¹²⁴ and sentence lengths (again). Despite the energy that was expended on the effort, however, these researchers—usually aided by the critiques of their interlocutors—discursively discovered that none of their efforts using any single measure—parametric or otherwise—seemed to consistently discriminate between authors.¹²⁵ The manifest failure to discover a single “silver bullet” for authorship begs the question...why? Five answers can be submitted for consideration.

First, the process of summarization itself implicit in the creation of any distributional measure can never, at least in Hoover's view, *ever* get us to “the silver bullet.”^[a] While Hoover's

¹¹⁵ Baayen termed this the “quest for characteristic constants.” Baayen, *Word Frequency Distributions*, 24–31.

¹¹⁶ Yule, “On Sentence-Length as a Statistical Characteristic of Style,” *passim*.

¹¹⁷ See especially Thomson and Thompson, “Outlines of a Method,” 59–64; Chotlos, “A Statistical and Comparative Analysis of Individual Written Language Samples,” 86–94.

¹¹⁸ Simpson, “Measurement of Diversity,” 688.

¹¹⁹ The Sichel distribution (Sichel, “On a Distribution Representing Sentence-Length,” 25–34) was used as fodder by Radday et al., “Genesis, Wellhausen and the Computer,” 478; Radday and Shore, *Genesis: An Authorship Study*, 195–202; Holmes and Forsyth, “The Federalist Revisited,” 114–18; Gani, “Characterizing an Author's Vocabulary,” 1–10 and Gurney and Gurney, “Authorship Attribution of the *Scriptores Historiae Augustae*,” 126.

¹²⁰ Greenberg, “The Measurement of Linguistic Diversity,” 109–15; Herdan, *Type-Token Mathematics*, 92–95.

¹²¹ See especially Thoiron, “Diversity Index and Entropy” 199–200; Fucks, “On Mathematical Analysis of Style,” 127–28; Carroll, “Vectors of Prose Style,” 287; Radday, “Isaiah and the Computer,” 70–71.

¹²² Radday and Pollatschek, in particular, explored the work of Guiraud (Guiraud, *Les caractères statistiques du vocabulaire*, 68–71) in their work on vocabulary richness in the OT. See Radday and Pollatschek, “Vocabulary Richness,” 340.

¹²³ See specially the review of Edmundson, “Statistical Inference,” 116–18.

¹²⁴ Examples include: Mascol, “Curves of Pauline and Pseudo-Pauline Style I,” *passim*; Mascol, “Curves of Pauline and Pseudo-Pauline Style II,” *passim*; Chaski, “Empirical Evaluations of Language-Based Author Identification Techniques,” 4–41; O'Donnell, “Stephen Crane's the O'Ruddy,” 110; Jin and Murakami, “Authors' Characteristic Writing Styles,” 63–76.

¹²⁵ See Yule, “On Sentence-Length as a Statistical Characteristic of Style,” 370; Williams, “A Note on the Statistical Analysis of Sentence-Length,” 357–61; Wake, “Sentence-Length Distributions of Greek Authors,” 331–46.

concern was vocabulary richness, his summary applies equally well to any omnibus summary measure:

Despite the attractiveness of vocabulary measures... such measures cannot provide a consistent, reliable, or satisfactory means of identifying an author or describing a style. There is so much intratextual and intertextual variation among texts and authors...[such measures]...should be used with great caution, if at all.... Unfortunately, the long-cherished goal of a measure of vocabulary richness that characterizes authors and their styles appears to be unattainable. The basic assumption that underlies it is false.¹²⁶

Second, a hint may be provided by Brainerd. While attempting to find a best fit model of the type/token ratio, he discovered very different upper limits on Shakespeare's vocabulary for comedies, histories, and tragedies. Brainerd concluded, correctly in our view, that "a search for an explanation of this phenomenon might prove very rewarding."¹²⁷ Given that SFL theory has long viewed distinct genres/registers as "variation according to use," different genres/registers are necessarily defined in terms of a culturally shared down-selection—a pruning of the choices available to the speaker or writer. This pruning would include, of course, lexical diversity. Accordingly, a measure that attempts to solve for authorship without taking into account the variation caused by any components of sociolect such as genre/register becomes an exercise that ignores a major moment of variation in the sample. Initial verification of this theoretical hypothesis can be inspected in the work of Stamatatos et al., who discriminated between text *genres* successfully using the most frequent words in texts.¹²⁸

Third, Wake's work on sentence lengths, in which he claims excellent authorial separation, is a case in point. A closer reading of his data demonstrates that (1) when the more

¹²⁶ Hoover, "Another Perspective on Vocabulary Richness," 173. Cf. Tweedie and Baayen, "How Variable May a Constant Be," 331.

¹²⁷ Brainerd, "On the Relation between Types and Tokens," 518. Brainerd's mild critique of Gani (Gani, "Some Stochastic Models," 313–24) followed the same reasoning. Brainerd, "Some Elaborations upon Gani's Model," 459.

¹²⁸ Stamatatos et al., "Text Genre Detection Using Common Word Frequencies," 809–13.

standard level of significance testing is utilized ($SE \times 1.96$) the disputed *Historia* is actually *not* significantly different from his zoological works¹²⁹ and (2) as Herdan would later note, Wake's own data showed that the works by Xenophon and Aristotle were written by the same author!¹³⁰

Fourth, perhaps the entire exercise was simply doomed from the start. In this view the notion of a silver bullet itself has major flaws: (1) It underestimates the linguistic diversity of language-in-use in all its paradigmatic, syntagmatic, stratal, and functional potential.^[b] (2) It operates almost solely at the level of lexis. (3) It simplifies the nested hierarchies and participant chaining of coherent discourse structure to collocations. (4) It unitizes the multidimensionality of language that serves as a container for its variance and covariant structure. (5) It dechronologizes language and the information flow encoded in discourse development. In sum, linguistic silver bullets likely fail because they fail *linguistically*.

Fifth, and finally, why do ubiquitous function words (particles, conjunction, articles and the like) seem to consistently discriminate authorship better than any other measure? The reason for this, we submit, is very similar. Vectors of the most frequent words or especially the most frequent function words retain in some cases sixty to seventy percent of the noncontextual tokens in the text. The implication? Function words, or high-frequency words, by their very ubiquity, represent, quite accidentally, a balanced set of the systems of the language. Said more memorably perhaps, high-frequency noncontextual words discriminate between authors best precisely because they are noncontextual and because they have managed to linguistically exanguinate our hapless corpus the *least*.¹³¹

¹²⁹ To the best of my knowledge this fact has not yet been published. Using Wake's own data, given that the mean and standard error for his Zoological works are 18.4 and 0.8; the lower bound is $18.4 - (1.96 \times 0.8) = 16.872$. Moreover, given that the mean and standard error for the *Historia* are 15.7 and 1.6; the upper bound is $15.7 + (1.96 \times 1.6) = 18.836$.

¹³⁰ Herdan, *Type-Token Mathematics*, 57.

¹³¹ Using another metaphor, again from biology, language analysis is least destructive of the structure of the text when we study language *in vivo* rather than *in vitro*. Just as the ecologist or biologist learns more about how his

3.6.3 Design Implications

The implication for our experimental design is to do essentially the opposite of what this period attempted to do. It is to: (1) resist using a single distributional measure to attempt to summarize the distribution,^[c] (2) study the language of the text *in situ*—to study it in terms of its paradigmatic systems, syntagmatics ranks, stratal layers, and multiple metafunctions, (3) explore the nested hierarchies of coherent discourse structure^[d] by annotating that structure in terms of word groups, clauses, and clause complexes, (4) retain the multidimensionality of language by studying its variance and covariant structure and exploring the possibilities of unobserved latent structure^[e] and (5) explore the linear information flow^[f] encoded in discourse development.

3.7 Milestone Three: First Light: Mosteller, Wallace and Herdan (1950s–1966)

3.7.1 Mosteller and Wallace and Classical and Bayesian Statistics

Mosteller and Wallace were, and to a great extent still are, the towering figures in mid-twentieth-century stylometry. Their quarry was the authorship of the Federalist papers and there is scarcely a stylometric/computational stylistics survey that fails to mention their contribution glowingly.¹³² What did they do right? The answer is—*much*. Said in its briefest compass, for the first time in stylometric praxis, Mosteller and Wallace combined six critical success factors within one study.

object of study functions by observing it in the wild than shooting it and taking it back for dissection, so too language should be studied in its natural language-in-use habitat, in the dynamics of its instantiations, and in the messiness of its idiolectic and sociolectic interactions.

¹³² In the arena of computational stylistics, Mosteller and Wallace teamed first on their JASA paper (Mosteller and Wallace, "Inference in an Authorship Problem," 275–309) and then a year later Mosteller published essentially the same material in the Addison-Wesley series in behavioral sciences edited by Wallace (Mosteller and Wallace, *Inference and Disputed Authorship*). Fully twenty-one and twenty-four years later, Mosteller would update his initial work with some new lexically derived discriminators. See Mosteller, "The Writing Styles of the Authors of 'the Federalist' Papers," passim; Mosteller, *Applied Bayesian and Classical Inference*, passim. Lastly, forty-six years after their initial work, Mosteller's estate published his posthumous autobiography, Mosteller, "Who Wrote the Disputed Federalist Papers," passim.

First, they chose an *ideal authorship attribution* (IAA) task,¹³³ one with large numbers of known texts from separate authors. Second, they tested a wide variety of different *classes* of linguistic measures,^[a] not just a few *ad hoc* measures. Third, they tested these distinct classes of measures separately,^[b] not mixed together in a grab bag. Fourth, they used multiple independent quantitative methods to perform the prediction or discrimination between them.^[c] Fifth, they did what all good statisticians do; they handcrafted their mathematical model to maximize some measure of objective fit. Sixth, their validation method employed multiple, independent methods of inference: both Fisherian inference (which they referred to as classical inference) and Bayesian inference,¹³⁴ an alternative, distribution-free approach to significance testing.^[d]

And what did they find? Mosteller and Wallace's conclusions sent shock waves throughout stylometry: they determined that all but perhaps one of the contested Federalist papers were written by Madison.¹³⁵ For our purposes, however, this is almost beside the point. At least nine key experimental design findings can be drawn from Mosteller and Wallace's study. These will constitute our implications as well.

3.7.2 Design Implications from Mosteller and Wallace¹³⁶

1. Use word (lexical-level) features but do not limit the experimental design to the rank of lexis.^{137[e]}

¹³³ See Section 2.3 or the glossary for a further definition of IAA.

¹³⁴ Bayesian approaches to inference actually predated classical Fisherian statistics chronologically but they were eclipsed by classical approaches until their revival in the late twentieth century. For an accessible introduction to Bayesian inference see Tiao, "Inference and Disputed Authorship," 306–08.

¹³⁵ Mosteller and Wallace, "Inference in an Authorship Problem," 306.

¹³⁶ See also Sarndal's evaluation (Sarndal, "On Deciding Cases of Disputed Authorship," 251–55).

¹³⁷ See Mosteller and Wallace, "Inference in an Authorship Problem," 277. This underscores that with some serendipity involved, a single word may idiosyncratically distinguish one author from another, especially in IAA-type tasks. This strategy, however, is far less likely to work when trying to distinguish one author from many others.

2. Recognize that lexical choice is topical, not merely authorial; the rate of certain words “depends on the topic under discussion.”¹³⁸ Certain lexical stocks are associated with certain cultural and situational contexts. Test for these alternatives.^[f]
3. Use a wide variety of *types* (categories of the measure being used, such as a large set of lemmas within the set of all lemmas) not just a few *ad hoc* measures.^[g] Test distinct classes of measures separately, not mixed together in a grab bag.^[h]
4. Execute data cleaning prior to modeling. Only allow the best discriminators into the model.^{139[i]}
5. Handcraft the mathematical model to maximize some measure of objective fit.¹⁴⁰ A good model is one that fits the data.¹⁴¹
6. Implement supervised mathematical models if the task requires classification (i.e. discriminant analysis).^{142[j]}
7. Recognize that the best predictors are not low-frequency *hapaxes*¹⁴³ but high-frequency function words.^{144[k]}
8. Use multiple, independent methods of inference.^[l]

In sum, Mosteller and Wallace’s work was simply a remarkable achievement for its era.

It stands as a landmark for careful experimental design and conservative interpretation. They

¹³⁸ Mosteller and Wallace, “Inference in an Authorship Problem,” 278.

¹³⁹ Mosteller and Wallace, “Inference in an Authorship Problem,” 282–85.

¹⁴⁰ Mosteller and Wallace, “Inference in an Authorship Problem,” 279.

¹⁴¹ The main models used by Mosteller and Wallace were a linear discriminant model and a number of Bayesian model forms; Mosteller and Wallace, “Inference in an Authorship Problem,” 289–97.

¹⁴² Mosteller and Wallace, “Inference in an Authorship Problem,” 281–82. For explication of linear discriminant analysis see Section 3.9.3.

¹⁴³ Mosteller and Wallace minced no words in writing, “... we eliminated words whose low frequency (five occurrences or fewer) make them nearly useless for discrimination.” Mosteller and Wallace, “Inference in an Authorship Problem,” 280. See also Sarndal, “On Deciding Cases of Disputed Authorship,” 252. For the continuing mistaken overestimate of *hapaxes* as superior authorial discrimination see Labbé, “Authorship Attribution by Intertextual Distance,” 48.

¹⁴⁴ Mosteller and Wallace, “Inference in an Authorship Problem,” 283–85.

anticipated developments that would become commonplace decades into the future: multiple linguistic measures, multiple methods of analytical inference (both classical and Bayesian), the use of function words (The “Burrows’ Method”)¹⁴⁵ and the principled reduction of both Type I and Type II error. Just as importantly they recognized *ad hoc* discriminators as analytically unfruitful.¹⁴⁶ Mosteller and Wallace spawned a handful of scholars who followed profitably in their wake, including Sarndal who offers a classical statistical alternative to Bayesian hypothesis testing, yet without invalidating Mosteller and Wallace’s approach.¹⁴⁷

3.7.3 The Linguist-Statistician Gustav Herdan

Gustav Herdan, like Fisher before him, seemed to be a scholar without boundaries. His first doctorate was in law (1923 from the German Charles-Ferdinand University in Prague), followed by a PhD in Chinese Studies and English Philology (1937, Vienna). Escaping Germany in 1938, the then forty-one-year-old scholar emigrated to England and there studied mathematics and statistics, eventually becoming a lecturer in statistics at the University of Bristol. In Herdan, then, we find in one archetype advanced linguistic and mathematical training combined with the discursive capabilities of a lawyer.

3.7.4 Design Implications from Herdan

Herdan was such a central figure, and his work so exemplary, that four themes that characterized his professional life can constitute design implications for our own work:

Representativeness.^[m] In Herdan’s day the ability to develop *cell-level* nonparametric models of language-in-use, models that included every type (category) of a given language

¹⁴⁵ See especially Forsyth et al., “Investigating the Authenticity of the Consolatio,” 380; Juola et al., “A Prototype for Authorship Attribution Studies,” 173; Drew and Craig, “Did Dickens Write ‘Temperate Temperance’,” 267.

¹⁴⁶ In the words of Mosteller and Wallace the hand-picked markers were “annihilated” by machine-picked markers (Mosteller and Wallace, “Inference in an Authorship Problem,” 285).

¹⁴⁷ Sarndal, “On Deciding Cases of Disputed Authorship,” 264–67.

measure (e.g. all the words used in a corpus like the GNT) simply did not exist.¹⁴⁸ Accordingly, Herdan did the next best thing. He created summary statistics that represented those same underlying distributions, hence his C-quantity formula and the like. This places Herdan directly in the heritage of Yule who, while handicapped by the same limitation, nonetheless sensed that if one cannot build individual terms in a model that represented types then at the very least they can be embedded within a single, scalar summary statistic.

Linguistic sensitivity. Herdan displayed a linguist's sensitivities.¹⁴⁹ This was a prescient achievement given that the 1950s and 1960s were relatively early in the subsequent evolution of language theory, and the paradigms of traditional grammar were still dominant.¹⁵⁰ Moreover, complicating this, the filial generation of the nineteenth-century comparative philologists was still dominant in many circles hawking its largely universalistic and Indo-European centric wares.¹⁵¹ These dual factors ensured that linguistic analysis remained disproportionately lexical, that syntagmatics only meant "syntax" and that discourse flow meant only "word order"—the SVO construct of traditional grammar. Against these tides Herdan attempted to explore language more linguistically,¹⁵² and to explain his findings hierarchically (via his famous discussion of

¹⁴⁸ In the modern era we have varieties of nonparametric multivariate analysis to perform this task. Our work in this regard can be inspected in sections 7.2-7.4.

¹⁴⁹ Herdan even notes the transcendence of language complexity above that which can be captured in information theory. Herdan writes, "Information theory as the quantitative expression of de Saussure's attempt to conceive of language as a code...has, no doubt, contributed to our realizing the difference between...objects in space and time...and [those]...of linguistic expression, but it has not done so completely." Herdan, *Type-Token Mathematics*, 20. Had Roman Jakobson displayed a similarly nuanced view of the ability of simple Information Theory to capture language complexity, the subsequent trajectory of twentieth-century linguistics may have been very different.

¹⁵⁰ Discourse analysis and functional grammar, for instance, were at best only embryonically present during this period. Herdan's view of discourse analysis embraced the early perspectives of Harris and Longacre (Herdan, *Type-Token Mathematics*, 59-69) but Herdan died in 1966, long before corpus linguistics would begin its campaigns to experimentally inform discourse analysis with actual data.

¹⁵¹ Mention needs be made of Greenberg's work. His careful study demonstrated evolution in languages, and the extent to which the linear flow of information in *Indo-European* languages is similar, and pragmatically constrained. It does, however, say little about the breadth of that universality across all language groups in general. Greenberg, "Some Universals of Grammar," 73-113.

¹⁵² Evidence of this is found especially in his foray into semiology and his concerns with phonemes in Chapters III and VIII in his *Type-Token* monograph (Herdan, *Type-Token Mathematics*, 70-82, 127-33).

type/token/text and type/token/text/content).¹⁵³ Moreover, Herdan viewed the phenomenon of choice in the “remaining features of language”¹⁵⁴ as central, as does the modern systemic functional linguist. Lastly, Herdan localized choice to those parts of language neither above the control of the individual (e.g. genre or other sociologically reduced choices in the system networks of the speech community) nor below the control of the individual (the phonemic).¹⁵⁵

His Inductivist and Descriptivist Sense. Herdan’s theories were thoroughly inductivist. Like Boas before him, the instances of parole *mattered* to Herdan. As a statistician, however, he did what statisticians *do*—he disemboweled their linear flow, aggregated those instances, and studied the various parameters of those distributions. Despite this, however, Herdan’s quantitative inductivism was salutary for two reasons. First, it led him to test various scalar or distribution-based linguistic measures such as sentence lengths and Zipf’s Law, only to discover them wanting.¹⁵⁶ Second, it enabled him to discover repeated patterns in the distributions from different authors. These he reduced to symbolic mathematics which he—in the practice of the day—expressed as “laws.” These for Herdan included the law of relative growth of literary products, the type token ratio,¹⁵⁷ and, in collaboration with Grayston, the C-quantity formula. The importance of the latter, it should not be missed, lay in the fact that C was a very necessary correction to Harrison’s work, one that normalized *hapax legomena* by text length.

Limitations of Herdan. Herdan’s works, of course, do not lie beyond critique. A pervasive structuralism, which few thought odd at the time, runs through his work.¹⁵⁸ This

¹⁵³ Herdan, *Type-Token Mathematics*, 243–44, 55–61.

¹⁵⁴ Herdan, *The Advanced Theory of Language as Choice and Chance*, 5.

¹⁵⁵ Herdan, *The Advanced Theory of Language as Choice and Chance*, 4–5.

¹⁵⁶ Herdan, *Type-Token Mathematics*, 56; Herdan, *The Advanced Theory of Language as Choice and Chance*, 88.

¹⁵⁷ Herdan, *Type-Token Mathematics*, 83–91.

¹⁵⁸ Indeed Herdan viewed his task as liberationist, “to apply statistical procedures to the problems of structural linguistics, thus raising this branch of linguistics to the level of a science.” Herdan, *The Advanced Theory of Language as Choice and Chance*, 4.

explains how Herdan, a linguist who appreciated the multilayered complexity of language, nonetheless imagined, along with other structuralists, that it could be meaningfully modeled by simple single channel information theoretic approaches.¹⁵⁹ It is evident too, that Herdan built his quantitative approach upon a Saussurean backbone – which served as a launching pad for his intriguing (but ultimately unconvincing) duality conception. This, too, was a generalization of linguistic structuralism, which he termed *literary statistics*.¹⁶⁰ In so doing he accepted uncritically the *langue/parole* distinction, and other tenuous structuralist distinctives.¹⁶¹

3.7.5 Herdan and the GNT

Last, Herdan, similar to other researchers of his time, affirmed multiple causes of linguistic variation,¹⁶² but when actually faced with vocabulary distribution inhomogeneity by text grouping (in the GNT no less) he nonetheless defaulted to the *de facto* assumption that its cause lay in authorship.¹⁶³ Regarding the latter, in Grayston and Herdan’s “Authorship of the Pastorals in the Light of Statistical Linguistics” for instance, Herdan opposes Guthrie’s arguments against P.N. Harrison by affirming the truism that *speech communities* (Herdan’s expression) display similarity not only in choosing characteristic linguistic features but also in

¹⁵⁹ It should be noted that even Jakobson was seduced by the same intellectual ambush. Chomsky’s *Syntactic Structures*, which fell in between the publication of Herdan’s major works, demolished the structuralist optimism that language could be modeled so simplistically (Chomsky, *Syntactic Structures*, 19–21).

¹⁶⁰ Herdan put much stock in *literary statistics*; for him it was the verificational completion of the incomplete program of linguistic structuralism, and “as a quantitative philosophy of language should apply to all branches of linguistics.” Herdan, *The Advanced Theory of Language as Choice and Chance*, 9.

¹⁶¹ Herdan, *Type-Token Mathematics*, 230; Herdan, *The Advanced Theory of Language as Choice and Chance*, 7.

¹⁶² Herdan clearly viewed “subject matter, and chronological distance” as sources of variation (Herdan, *The Advanced Theory of Language as Choice and Chance*, 257) as well as literary genre – at least at the coarsest level, the difference between poetry and prose (Herdan, *Type-Token Mathematics*, 260; Herdan, *The Advanced Theory of Language as Choice and Chance*, 206–13).

¹⁶³ Herdan concludes that the proposition that a single Paul wrote the 13 NT texts traditionally ascribed to him as “doubtful.” See Herdan, *Type-Token Mathematics*, 250. Here, however, Herdan’s characteristic caution in assigning causation abandons him. Underlying this conclusion lay his assumption that the largest component contributing stylistic variability is authorship, a fact that Herdan accepts but does not demonstrate. Adding to the inconsistency Herdan later says that the variability of Herdan’s V within a series is due to style and content (Herdan, *Type-Token Mathematics*, 252).

grammaticalizing them at characteristic frequencies.¹⁶⁴ But Herdan neglected to appreciate that his truism only followed *ceteris paribus*, given that other sociocultural components of variation (genre, date, audience and the like) are held constant.¹⁶⁵ Last, Herdan mercilessly deconstructed the four errors of Harrison,¹⁶⁶ covering that ground with salt, but oddly reproduced Harrison's most fundamental error, assigning significant variation to authorship without quantitatively eliminating other potential causes for that variation.¹⁶⁷

3.8 Period Four: Stylometry Amidst the Positivist Bacchanal (1930s–1980)

In the period between 1930 and 1970, it seemed that all the necessary stylometric pieces were now in place to solve some of the more vexing problems of the discipline. Stylometric workers now had at their disposal new categories of linguistic measures from general/theoretical linguistics, an early though solid set of multivariate analytics,¹⁶⁸ and *two* flavors of mathematical inference from which to draw (Bayesian and Classical). Conceptually, one might conclude, nothing prevented researchers using these methods from yielding new stylometric insights all harnessed under a careful sequential experimental design. Inspection of the literature of the period, however, demonstrates that most mid-twentieth century research neither attained to this lofty level of integration, nor even to the standards set by Mosteller and Wallace. Rather, most

¹⁶⁴ Grayston and Herdan, "Authorship of the Pastorals," 2–3.

¹⁶⁵ This is clearly a corpus-level statement, and while it clearly demonstrated Grayston and Herdan's prescience, no adequate sized corpora existed in their era to verify it.

¹⁶⁶ Grayston and Herdan, "Authorship of the Pastorals," 4–7.

¹⁶⁷ Grayston and Herdan, "Authorship of the Pastorals," 10–15. In a rare inconsistency in any paper written by Herdan, the authors on page 10 assert that "the magnitude of C... supports strongly the hypothesis of a non-Pauline authorship," and then Herdan retreated from that position by concluding on page 15 that "whether this implies a difference in authorship depends on one's conception of... style."

¹⁶⁸ By the mid-twentieth century the parametric arm of multivariate tools included principle components analysis courtesy of Hotelling, "Analysis of a Complex of Statistical Variables," 498–520, factor analysis (Thurstone, *The Vectors of Mind*, 1–133), multiple discriminant analysis (Fisher, "The Use of Multiple Measurements in Taxonomic Problems," 179–88; Tatsuoka and Tiedeman, "Discriminant Analysis," 402–20) and analysis of variance in its many forms as originally developed by Gauss and Laplace and formalized experimentally by Fisher in 1918 (Fisher, "The Correlation between Relatives," 399–433; Fisher, "On the 'Probable Error'," 3–32). The nonparametric arm of the multivariate revolution would emerge latter in this period under Nelder and Wedderburn (Nelder and Wedderburn, "Generalized Linear Models," 370–84; Nelder, "Log Linear Models for Contingency Tables:," 323–29).

research in this period was executed *without* those emergent linguistics measures, using univariate rather than multivariate analysis, and was verified with only rudimentary tools of classical inference. The question this raises, again, is *why*. To a certain extent it is always speculative to try to distill from history why a given period missed its own opportunities. However, at least two things seem clear. First, this period of stylometry emerged at precisely the same time two of the greatest rivers in twentieth-century intellectual thought crested: structuralism and positivism. The latter was an epistemological stance that proposed, baldly, that every necessary truth is only known *a priori* and every contingent truth is only known *a posteriori*.¹⁶⁹ Its limiting of prior knowledge to logic and posterior knowledge to empiricism radically altered the stance of the logical positivist researchers not only toward the rest of the academy but also toward their own objects of study. The stylometric researchers who embraced this epistemology either explicitly or pragmatically did their science (for so they termed it) with an epistemic hegemony that had never before or since been known. Understanding these forces and their influences on stylometry are the tasks we will explore next.

3.8.1 Late Nineteenth Through Mid-Twentieth Century Stylometric Positivism

The trajectory of Logical Positivism is generally well understood: its inspiration by Comte,¹⁷⁰ its birth in the Berlin and Vienna circles, its peak under Wittgenstein and Russell,¹⁷¹ its merger with Otto Neurath's physicalism and Carnap's reduction of philosophy to science,¹⁷² its

¹⁶⁹ McLeod, "Knowledge of Necessity," 179.

¹⁷⁰ See especially Comte's six volume manifesto; Comte, *Cours de philosophie positive*.

¹⁷¹ For the collective pessimism of Wittgenstein, Russell and Neurath on the meaning-content of non-empirical propositions see Wittgenstein, *Tractatus Logico-Philosophicus*, Proposition 4.003; Russell, "On Denoting," 479–93; Neurath, "Physicalism," 619. For Whitehead's and Russell's view of the intra-systemic consistency of mathematics, see Whitehead and Russell, *Principia Mathematica*, vii–viii.

¹⁷² See Neurath, "Physicalism," 618–23; Carnap, "On the Character of Philosophic Problems," 506.

mathematical unraveling under Gödel,¹⁷³ and its philosophical inadequacy as discursively demonstrated by Quine, Popper, and Kuhn.¹⁷⁴

While this work's scope cannot permit delving into positivism's influence over stylometry in detail,¹⁷⁵ it would be a mistake to not highlight a few of its central effects especially as they peaked in the late nineteenth through the early twentieth centuries. First, it should be noted that in those halcyon days before indeterminacy in its many forms carried away positivism's collective epistemological innocence, an expansive flavor of populist positivism intertwined itself with stylometry's earliest efforts. We have seen this already in Moritz, who approvingly quoted Lagrange's wholly mechanistic view that "the complete history of the universe could be represented by one huge differential equation."¹⁷⁶ Yet, as the positivist scientific enterprise gathered steam in the early half of the twentieth century, and structuralism along with it, it constituted a kind of perfect storm. Unbridled scientific achievement, technological optimism and naïve empiricism created an ethos in which the market in fallacies of composition was never hotter, the frequency of epiphanal findings never higher, and the linear distance between observations and conclusions never shorter. It was, simply, the *force majeure* of its time—and many empiricists, not just stylometricists—imbibed at its bacchanal.¹⁷⁷ The

¹⁷³ See especially Gödel's fourth and eleventh theorems; Gödel, "Über formal unentscheidbare Sätze der Principia Mathematica," 180–86, 96–98.

¹⁷⁴ Quine rejected the radical reductionism of Carnap (Quine, "Two Dogmas of Empiricism," 37–39) while Popper critiqued all justificationalist systems including both logic and empiricism (Popper, *Conjectures and Refutations*, 342–76). Kuhn perhaps more than any other postpositivist thinker relativized positivist claims through his notion of incommensurability; Kuhn, *The Structure of Scientific Revolutions*, 147.

¹⁷⁵ See the reviews of Robins and Morgan for the influence of positivism on linguistics; Robins, *A Short History of Linguistics*, 236–54; Morgan, "Poststructuralism and Applied Linguistics," 1035–38. For the reaction against positivism in literary theory, see Hirsch, *Validity in Interpretation*, 145; Bakhtin, *Speech Genres and Other Late Essays*, 141. For the pragmatic turn in postpositivist science after the collapse of logical positivism see Horowitz, "Two Cultures of Science," 437; Michell, "Pragmatism, Positivism and the Quantitative Imperative," 43–52.

¹⁷⁶ Moritz, "On the Variation of Certain Sentence-Constants," 229.

¹⁷⁷ Yet, even as positivism peaked, many empiricists did resist the hegemony and rush to conclusions. In her response to Milic, Matlack demonstrates the reserve that breathes at the core of all careful empirical work as she writes, "We must first of all recognize that a scholar does not approach the data with a completely open mind, but is

result? Many studies in this period, unaware of, or underestimating the linguistic complexities of actual language in use, became serial offenders in the commission of Type I *causal* error: the assigning a cause to effect A when in fact that cause was due (or principally due) to one or more other causes.¹⁷⁸ The task I suggest, then, in reviewing this period is to chart a mediating course, to neither overestimate (Type I error) nor underestimate (Type II error) the influence of positivism on mid-twentieth-century stylometry. To do so this review will proceed in two stages. First, this period will be summarized using its three most storied examples in biblical studies, A.Q. Morton, Yehuda Radday, and Ronald E. Bee. Second, the linguistic, analytic and experimental design choices of these scholars will be explored, in order to draw from them principles useful for the experimental design to be defined and articulated in Chapter Four.

3.8.2 The *Paterfamilias* of Classical and Biblical Stylometry: A. Q. Morton

It is seldom advisable to choose a single individual to represent an entire era in any discipline. Yet, if we were to do so in stylometry, that period would be the mid-twentieth century and that person would be Andrew Queen Morton. Morton and his methods have been subjected to a number of grisly academic postmortems, of course, notably the critiques of Deane against sentence lengths,¹⁷⁹ Damerau against function words,¹⁸⁰ Weitzman against overeager assignment of plural authorship,¹⁸¹ and Hilton and Holmes against QSUM charts.¹⁸² Yet, we miss the nexus

armed with intuition, and in some cases, with external evidence.” See Matlack and Matlack, “A Statistical Approach to Problems of Attribution,” 629.

¹⁷⁸ This is also the commission of Type II error because the actual cause was missed.

¹⁷⁹ Deane, “Stylometrics Do Not Exclude the Seventh Letter,” 113–17.

¹⁸⁰ Damerau objected to the idea of Morton’s use of certain function words as a universal (Damerau, “The Use of Function Word Frequencies,” 272–74). Cf. also Oakman, *Computer Methods for Literary Research*, 142.

¹⁸¹ Weitzman’s observations are found in the response to Bartholomew (Bartholomew, “Probability, Statistics and Theology,” 173). Cf. again to Oakman’s assessment of Morton’s work as substantially incautious (Oakman, *Computer Methods for Literary Research*, 143–46).

¹⁸² For Morton’s use of QSUM charts see Morton and Winspear, *It’s Greek to the Computer*, 80–82; Morton, *Literary Detection*, 77–92. For Hilton and Holmes’ critique, see Hilton and Holmes, “An Assessment of Cumulative Sum Charts,” 74–79. See also Holmes summary of the decline and fall of Morton’s methods in his review articles (Holmes, “Evolution of Stylometry,” 113; Holmes and Kardos, “Who Was the Author,” 7–8).

of the critiques against Morton if we fail to listen to Morton's two most stalwart opponents, Gustav Herdan and M.W.A. Smith.¹⁸³ Both researchers, in successive eras, would come to spare no quarter in deconstructing Morton's linguistic measures as well as his scientific praxis.¹⁸⁴ If we were to synthesize their separate critiques, we would conclude that even if some sort of extraordinary indulgence were granted Morton for being a pioneer just as positivism ran its hottest, two factors place him in an entirely separate category from other scholars reviewed here. First, as one who did empirical science he utterly failed in doing the only thing scientists must do: be as large in their experimental design as they are cautious and thorough in their conclusions. Morton did just the opposite. This bore especially bitter fruit for New Testament scholarship, because as a theologian he earned a platform few do—he stood in front of the Royal Statistical Society on that long-ago Wednesday in January of 1965 to present a paper on stylometry, including findings related to the GNT. On that day, and regrettably subsequently, he failed to do good science and was pilloried for it. Second, he failed to retract his views even as positivism imploded around him, and a chastised, humbler form of empirical science eventually emerged. The former failure was in a sense recoverable and excusable, the latter was not. We will treat Morton's earlier and later work separately.

Morton's Early Work. In his inaugural paper, "The Authorship of Greek Prose," Morton chose sentence lengths as Wake had done before him and analyzed them with Yule's *K*.^[a] In the eight intervening years since Wake's paper however, sentence lengths had, deservedly, gone into eclipse (per our review of Wake in section 3.6.2). Yule (1939), moreover, had expressed multiple

¹⁸³ See especially Herdan's response in Morton, "The Authorship of Greek Prose," 229–31 and Smith's latter work; Smith, "Computers, Statistics and Disputed Authorship," 376–95.

¹⁸⁴ Smith writes as late as 1994 regarding Morton, "From a scientific standpoint it has been disappointing to have demonstrated the scale of the inadequacy of documentation, the lack of systematic procedures for selecting data, the haphazard approach to statistical inference, and the limited number and incomplete nature of the tests for establishing the validity of the method" (Smith, "Computers, Statistics and Disputed Authorship," 394).

reservations regarding sentence lengths,¹⁸⁵ and Brown (1963) had classified them as among the least valuable criteria for authorship attribution.¹⁸⁶ When Morton omitted any discussion of these works, Herdan took the opportunity as a member of the Society to attach his comments to the paper. He issued an eightfold critique charging that Morton, among other things, “omitted...adequate documentation indispensable in a scientific paper,” “made...statements which are obviously untrue,” “omitted all references to previous weighty criticism of the methods he employs,” “persistently confuses style and authorship,”^[b] [issued] “20 pages of irrelevant analysis....to preface...disappointing results for the Paulines,” and lacked “knowledge of the relevant literature.”¹⁸⁷ Herdan’s overall assessment was that Morton’s paper was to be welcomed, “even though, as in this case, it can only serve to demonstrate *how not to do it*.”¹⁸⁸ If Herdan’s response was trenchant, he had adequate reason. Herdan’s work, especially his 1956 monograph, *Language as Choice and Chance*¹⁸⁹ was arguably the most seminal work covering stylometry of its time—and Morton showed no evidence that he had ever even read it. Worse yet, five years prior to Morton’s paper Herdan had deconstructed Wake’s work—upon which Morton’s was based—because Wake’s efforts lacked discrimination.¹⁹⁰ Thus, while arguably Wake and Morton did demonstrate that certain hand-selected authors could be differentiated using Yule’s K, both failed to test the performance of Yule’s K with more similar authors and

¹⁸⁵ Yule, “On Sentence-Length as a Statistical Characteristic of Style,” 363–71. See especially his discussion of the difficulties of determining the length of sentences on pages 363–364. Yule’s reservations are even more acute in relation to the Greek of the Classical periods down to the Koine of the GNT, given that this era did not use anything akin to our own punctuation marks and what did exist in the text (full-stops and colons) were generally far more rare than our own periods.

¹⁸⁶ Brown, *The Authentic Writings of Ignatius*, 139.

¹⁸⁷ Morton, “The Authorship of Greek Prose,” 229.

¹⁸⁸ The italics above, it needs be said, are Herdan’s. Herdan’s response can be found appended to Morton’s paper (Morton, “The Authorship of Greek Prose,” 229). Morton’s later book (1966) *Paul, the Man and the Myth* added little quantitatively to the paper he read at the British Statistical Society, except to underscore its theological implications (Morton and McLeman, *Paul, the Man and the Myth*, 101–34).

¹⁸⁹ Herdan, *Language as Choice and Chance*.

¹⁹⁰ Herdan, *Type-Token Mathematics*, 55–58.

failed to test alternative causal explanations other than authorship.¹⁹¹ In so doing, Wake and Morton missed the central conclusion that *could* have arisen from their work, namely, that sentence-lengths assayed by Yule's K or any single summary statistic tended to produce unacceptably high numbers of false negatives.^[c]

Morton's Later Work. The later Morton seems to have been immune to the weighty eightfold critique of Herdan, and largely ignored it.¹⁹² It cannot be our goal here to summarize the rest of Morton's long career but to focus on only that part relevant to our own goal of developing an experimental design capable of disentangling authorship from genre. Ten summary statements can be made regarding Morton's latter trajectory as a stylometrician. *First*, Morton did *not* fail to recognize that genre and other socio-cultural factors contribute to the variation in language.¹⁹³ He simply made no room for them in his experiments.^[d] In so doing, he made two *de facto* assumptions: (1) that the major components of variation were due to authorship and (2) that variation from other causes was essentially negligible.¹⁹⁴ By assuming what he should have demonstrated, Morton's causal conclusions became circular. *Second*, he continued his habit of omitting the relevant literature and even appeared to take credit for already extant innovations.¹⁹⁵ *Third*, against the tide of the times, Morton became reductionist, using less

¹⁹¹ Even Moritz did so over seventy years prior to Morton (Moritz, "On the Variation of Certain Sentence-Constants," 235–43).

¹⁹² In response to Herdan's charges Morton only obliquely apologized as follows, "Dr. Herdan must learn that though some omissions are the result of ignorance or of carelessness, other omissions are critical." Morton, "The Authorship of Greek Prose," 233.

¹⁹³ Morton's term for genre was literary form (Morton, *Literary Detection*, 38).

¹⁹⁴ Craig terms this one of the signal pitfalls of authorship attribution, "Assuming that an author cannot vary from his or her normal style" (Craig, "Stylistic Analysis and Authorship Studies," 287).

¹⁹⁵ Morton refers to his own work (Michaelson and Morton, "The New Stylometry," 90), but neither to Penick's prior conditional sentences and articular infinitives (Penick, "Paul's Epistles Compared with One Another," 67–69), nor the word clusters of Frautschi (Frautschi, "A Project for Author Discrimination," 15), nor the dependent clause length and the prevalence of paratactic clauses in Hartwell (Hartwell, "A Quantitative Approach to Thomas Hardy's Prose Style," 152–54).

linguistic variety rather than more.^{196[e]} *Fourth*, Morton continued his practice of making confident and unqualified overstatements, stating as stylometric axioms what never has been empirically demonstrated.¹⁹⁷ *Fifth*, he continued to make patently unlinguistic choices, rooted in traditional grammar alone, for his language measures.¹⁹⁸ Morton chose, for example, what is arguably the most linguistically uncertain syntagmatic concept possible—the unpunctuated sentence in the Koine of GNT.¹⁹⁹ *Sixth*, Morton continued to demonstrate an uncanny ability to select measures associated with sociocultural sources of variation (such as first and last words of speeches even across disparate contexts, settings, and audiences) and assign them to authorship.^{200[f]} *Seventh*, Morton continued to overestimate the syllogistic efficacy of referencing astronomical values reported by simple tests of inference.²⁰¹ *Eighth*, we discover Morton making

¹⁹⁶ Michaelson and Morton’s decision to reduce authorship discrimination to a single word, perhaps not surprisingly, evidenced some untoward effects such as concluding that Hebrews was more likely Pauline than Ephesians. Michaelson and Morton, “The New Stylometry,” 101–02. This is not to deny that certain idiolectic expressions with high discrimination exist, see especially Merriam, “An Experiment with the Federalist Papers,” 253–54. Merriam’s work, however, was an exercise in the same genre, for the same purpose, in the same period, and with only two authors.

¹⁹⁷ Thus he immediately makes the same confident statements regarding his new adopted ken (the position of once-occurring words) and without evidence attaches it to his favorite muse: authorship. Thus Morton flatly states, “It is the position of these words that enables one writer to be distinguished from another” (Morton, “Once. A Test of Authorship,” 1). Here too is found Morton’s claim that “The nearest approach to an ideal habit for...authorship...[is] ...using the particle *de* as one of the first three words in sentences.” Morton and Winspear, *It’s Greek to the Computer*, 65.

¹⁹⁸ Mitigating this, Morton began to embrace the idea that collocations of linguistic units within syntagms (natural units of language comprised of more than one unit of language of lower rank) were worthy of exploration.

¹⁹⁹ Morton, “Once. A Test of Authorship,” 1–3.

²⁰⁰ And so Morton concludes that the statistical difference yielded when the frequency of the uniqueness of the first and last words in a sentence are compared between the “four genuine epistles” and the rest of the possible Pauline corpus, demonstrates authorship. Morton, “Once. A Test of Authorship,” 5. In so doing, however, Morton fails to fully appreciate the existence of epistolary sub-genre. See Chapter Five.

²⁰¹ Even later in his career, we still find Morton writing, “...on the evidence of this one Table, the probability that Peele wrote a play of Shakespeare is less than one in forty million.” Morton, “Once. A Test of Authorship,” 8. In contrast, Smith thoroughly deconstructs Lake’s naïve trust in such numbers, and underscores the underlying prerequisites needed to prevent false positives (Smith, “The Revenger’s Tragedy,” *passim*). Compare this practice with the unease Mosteller has in the kinds of overstatements such numbers invite (Mosteller, “The Writing Styles of the Authors of ‘the Federalist’ Papers,” 139).

pedagogical oversimplifications that do justice to neither linguistic²⁰² nor statistical theory.²⁰³

Ninth, even the nonquantitatively trained among Morton's reviewers began to notice Morton's recalcitrant use of questionable markers, and lack of controls.²⁰⁴

3.8.3 Ronald E. Bee

Bee published a trilogy of stylometric articles in the early 1970s (1971-1973) focused upon the Hebrew Bible. He christened his method "cumulative sum charts," an adequate enough summary for a method which plotted a transform of a running total of verb forms²⁰⁵ against the total number of words in a text. Bee's innovation lay in the compelling *face validity*²⁰⁶ of his charts; segments of the chart with higher or lower slope were highly recognizable. Bee assigned this change in slope to authorship (written/scribal vs. prophetic/oral). Similar to Morton, however, both Bee's methodology and his conclusions displayed a number of deep design flaws and interpretive problems.

First, his cumulative sum charts were based only on a single language measure,^[g] a running frequency of Hebrew verbal forms. Second, alternative causes for the change in slope were neither proposed nor investigated but assigned *de facto* to authorship.^[h] Our own analysis

²⁰² Classic among these is Morton's statement that "Stylometry begins with the realization that language is a sequence in which there are only two parameters, the number of times a word occurs and the position in which it occurs" (Unsworth and Morton, "Mrs. Gaskell Anonymous," 28). No modern linguistic theory can concur with a claim this reductionist. Even if we omit the most empirical of modern language schools from consideration as too particularist (the systemic functional perspective), then language still displays at least a dozen features that virtually all schools would agree exist: length (rank), order (permutation), combination (colligations and collocations), paradigmatic choice (systems), semantic grouping or domains, taxis, transitivity/ergativity, modality, polarity, tense/aspect, and functional relations (nominalization, modification and the like).

²⁰³ Here we find Morton quite erroneously claiming, for instance, "if we were to look at individual word-forms then the only ones which are used often enough...are the high-frequency particles and connectives." Morton and Winspear, *It's Greek to the Computer*, 48.

²⁰⁴ Griffith, "It's Greek to the Computer," 162.

²⁰⁵ The transform is equal to the number of verbs up to that point in the text minus a quotient, where the quotient merely allows the number of verbs (plotted on the ordinate) to take up less vertical space.

²⁰⁶ Face validity according to Gravetter et al., is the simplest and least scientific measure of validity. It "concerns the superficial appearance or face value of a measurement procedure." Quite often measures with high face validity are poor overall measures of validity when compared to coherence-based measures such as validity by fit or validity by reconstitution. See Gravetter and Forzano, *Research Methods for the Behavioral Sciences*, 78–81.

of Bee's data, for instance, demonstrates that differences between high and low slope regions are highly associated with shifts in the proportion of Hebrew tenses used, particularly between the *wayyiqtol* and imperfect tenses.²⁰⁷ Given the aspectual nature of the Hebrew verbal system, these shifts in proportion seem to be far more parsimoniously explained as shifts in aspect, which in turn are typically reflective of either discourse shifts (storytelling) within a given author, or shifts in literary subgenre (such as the shift from narrative to direct speech). Third, the change in slope was not determined mathematically, but by "eye."²⁰⁸ Fourth, in almost all cases the sample sizes in the individual segments of Bee's cumulative sum chart were too small to validly apply statistical tests.^{209[i]} Fifth, not infrequently his statistics were invalidly developed.²¹⁰ Sixth, Bee interprets the significantly smaller variance (sum of squares) for words counted separately when joined by a *maqeph* as evidence of scribal authorship as opposed to oral tradition. This assumes, quite remarkably, exactly what needed to be demonstrated.²¹¹ Seventh, the greatest problem in Bee's work lay in the very nature of classical CUSUM charts themselves. Not only do CUSUM charts lack a statistical basis,²¹² they display deep *frangibility*, the undesirable property in which small changes in data present large changes in visual display.^[j] The reason for this property is

²⁰⁷ In Bee's lower slope regions, specifically 19:1-11, 19:16-20, and Exodus 21:21-35, the ratio of *wayyiqtol* to the imperfect is 1.64. In the remaining higher slope regions that same ratio is 0.81. Given that *wayyiqtol*, the consecutive preterite, is often found in narrative and the imperfect in divine speech or quoted direct speech, it seems that if Bee has discovered anything at all, he has found a coarse method for identifying discourse shifts in Hebrew prose.

²⁰⁸ Bee, "A Statistical Study of the Sinai Pericope," 407. Mitigating this, the sections Bee developed did indeed demonstrate good statistical fit to an ordinary least squares regression line.

²⁰⁹ Some segments contained as few as three verses with only fifteen verbs (Bee, "The Use of Statistical Methods in Old Testament Studies," 260).

²¹⁰ His sum of squares were not independently developed for each segment and hence their residuals were not normally distributed; Bee, "Statistical Methods in the Study of the Masoretic Text," 613.

²¹¹ If one assumes, as Bee does, that later scribal editing is evidenced by a smaller variance (lower sum of squares and hence better fit) from the regression line when the abscissa is obtained by counting words separately (without the *maqeph*), it is decidedly not a finding for scribal authorship when one finds a stretch of verses that display such a lower variance. This is simply empirically discovering an instance within the data for what one has already predefined. Other interpretations are not explored by Bee. Becking, while addressing Bee's dating model raised this same issue: "a) Should not the lack of a *maqeph* rather be explained by the grammatical model the Masoretes [used]?" See Becking, "Bee's Dating Formula," 102.

²¹² Hilton and Holmes, "An Assessment of Cumulative Sum Charts," 73; Tweedie, "A Statistical Investigation into the Provenance of De Doctrina Christiana," 39.

that CUSUM charts were adopted from the industrial design industry. As implemented there, engineers quite validly used CUSUM charts not to make casual determinations but to flag potential real-time state changes in manufacturing processes.²¹³ Stylometricians adopted this metaphor then replaced time on the horizontal (x-axis) with the length of a running text and whatever linguistic “events” they chose (typically verbs) on the y-axis. This was an unfortunate decision in that by so doing, CUSUM charts present *discrete data as continuous data*.²¹⁴ Because representing the data discontinuously, however, will never produce a chart of high face validity because it will never yield a negative slope, stylometricians then made yet another unfortunate choice. They solved the “going negative” problem by plotting an odd sort of rate (the number of target words minus a transform of the running number of words) on the y-axis against the running number of words on the x-axis. This is quite illegitimate graphically because this makes the y-axis no longer independent of the x-axis, a signal property assumed in an x-y chart whose axes are plotted orthogonally (at right angles) from one another. Accordingly, a classical CUSUM chart will happily continue to draw an increasingly downward slope in the data, as the x-axis ticks off more and more words in the text, when in fact no new data instances of the given linguistic measure (often verbs) have occurred.²¹⁵ For a technique that achieves much of what CUSUM charts attempted but failed to do, see Girón, et al.’s use of Bayesian-predicated change points in the authorship of *Tirant lo Blanc*.²¹⁶

²¹³ See especially Hawkins and Olwell, *Cumulative Sum Charts and Charting for Quality Improvement*, 14–16.

²¹⁴ E.g. Fig. 1 in Bee, “The Use of Statistical Methods in Old Testament Studies,” *passim*.

²¹⁵ Bee’s reduction of potential causes to authorship alone were normative of positivistic science and the dominance of univariate methods in the mid-twentieth century (Bee, “A Statistical Study of the Sinai Pericope,” 409). Accordingly he neither mentions nor explores whether other factors could better explain the significantly different slopes in his cumulative sum charts.

²¹⁶ Girón et al., “Bayesian Analysis of a Multinomial Sequence,” *passim*.

3.8.4 George K. Barr

I have some deep reservations about including both G.K. Barr and Yehuda Radday (who we next discuss) within a section that even includes the phrase “Positivist bacchanal.” The positivist label fits neither of these two serious scholars well. Yet, as mentioned in the introduction to this section, even excellent scholars can occasionally wield inadequate methods and so it is here. With respect to Barr, many of the same critiques that apply to Bee’s work apply to Barr’s. Barr published, by my count, ten articles in the *Irish Biblical Review* and *Literary and Linguistic Computing* between 1996 and 2004 as well as a T&T monograph. Barr, as Bee did, adopted a visual metaphor to convey sentence length distributions through a variant of the now familiar rubric of cumulative sum charts.²¹⁷ The question is not whether Barr has used CUSUM charts as did Morton or Bee (he did not; his quarry related to the scale of sentences), nor whether he found data that supports his thesis (he likely did). Neither is it whether long syntagmatic stretches may not be modestly efficient at discriminating authors (a finding I differentially support). Rather, we wonder whether other linguistic species drawn from other points on either the paradigmatic cline or his preferred syntagmatic cline will evince precisely the opposite conclusions.^{218[k]} Barr has taken a very thin piece of linguistic data and made it epigrammatic for the whole. Like all CUSUM predicated work, this is stylistic synecdoche, if the expression may be pardoned, on a massive scale.²¹⁹ Finally, seen from the framework of experimental design,

²¹⁷ See especially the following pages in Barr’s work: Barr, “Contrasts in Scale and Genre,” 16–25; Barr, “The Use of Cumulative Sum Graphs,” 103–11; Barr, “Literary Dependence in the New Testament Epistles,” 148–60; Barr, “Scalometry and the Dating of the New Testament Epistles,” 71–90; Barr, “A Computer Model for the Pauline Epistles,” 233–50; Barr, “Graphical Analysis of the Sentence Length Distribution Curve,” 378–87; Barr, “Interpolations, Pseudographs, and the New Testament Epistles,” 438–48; Barr, “The Impact of Scalometry,” 5–9; Barr, “Two Styles in the New Testament Epistles,” 237–46; Barr, *Scalometry and the Pauline Epistles*, 27–106.

²¹⁸ It should be underscored that Barr claims to have found something more fundamental than the differences between authorship and genre. Noting a bipartite high-scale, low-scale structure in all traditional works of Paul, he writes, “In 1 Corinthians 1-6, 2 Corinthians 10-13, Philippians, and 1 and 2 Thessalonians, however, there is a similar structure but without any significant change in genre.” Barr, “The Impact of Scalometry,” 3.

²¹⁹ Hardcastle is more pointed, “The CUSUM method is seen to be discredited and should not be accepted as providing reliable evidence of authorship of either the written or the spoken word.” Hardcastle, “CUSUM,” 129.

because both Barr's and Bee's work rested upon some variant of CUSUM charts, their individual data charts were derivative, and nondistributional in nature. Hence they were not capable, except in the aggregate, of being explored by tests of statistical significance.²²⁰

3.8.5 Yehuda Radday

From the standpoint of stylometry Yehuda Radday is a unique case. He is difficult to review because like Nebuchadnezzar's vision, in this scholar we find both experimental iron and experimental clay. As an organizing framework for these comments the analysis of Radday's stylometry will be separated into two distinct periods: the earlier Radday, including his work on Isaiah and Zechariah (1970-1975), and the latter Radday, including his work in Judges, vocabulary richness and Genesis (1975-1985).

The Early Radday. Radday's early stylometric work focused primarily on the extent to which biblical Hebrew can be used to assay authorship hypotheses in the *Tanakh*. Several conclusions can be drawn from that work. Linguistically, what Radday occasionally lacked in currency (even when he wrote, sentence, syllable, and phoneme lengths were questionable discriminators)²²¹ he made up for with linguistic marker breadth²²² and a linguistically-precocious view of language strata. Concerning the latter, Radday was one of the very first to measure what SFL linguists would term extra-lexicogrammatical data—the stratal phenomena of

²²⁰ Partitioning of textual data and the overlaying of texts upon one another may make this possible, but I have seen neither.

²²¹ Herdan's critique of Morton's sentence lengths (Morton, "The Authorship of Greek Prose," 229–31) had been published five years before Radday's inaugural SBL article on statistical linguistics (Radday, "Two Computerized Statistical-Linguistic Tests," *passim*), but Radday nonetheless still opted for summary, length-based language measures. In addition Radday uncritically accepted the chestnuts surrounding the ubiquity of idiolect, that it is "beyond conscious control" and the like; Radday, "Two Computerized Statistical-Linguistic Tests," 319; Radday and Wickman, *The Unity of Isaiah*, 38.

²²² This includes, for instance, parts of speech (Radday and Wickman, *The Unity of Isaiah*, 135–57; Radday and Wickmann, "Unity of Zechariah," 37–38), transitions (Radday and Wickman, *The Unity of Isaiah*, 158–91; Radday and Wickmann, "Unity of Zechariah," 39), and particles (Radday and Wickman, *The Unity of Isaiah*, 192–203; Radday and Wickmann, "Unity of Zechariah," 36–37).

semantic “groups”.^{223[l]} Unfortunately—and here we begin to find the clay—Radday somehow viewed semantics, almost universally recognized as a reflection of subject matter or related to genre, to be a marker for authorship.²²⁴ Analytically, a similar story is found here, as well; what Radday lacked in quantitative currency (in his early work there is no hint of multiple methods of multivariate inference) he made up for by being one of the first to employ Information Theory-based measures, albeit again on the surprising choice of syllables.^{225[m]} Moreover, in his last major work on Genesis, Radday became one of the earliest to employ multiple methods of decompositional multivariate analytics to address issues of authorship.²²⁶ Experimentally, the story is mixed here as well. Positively, from the very start, Radday was clearly sensitive to the need for careful experimental design. Yet, in his table of chi-squares of inflected nouns in Isaiah, for instance, he makes no mention of alternative sociolectic, dialectic or diachronic causes.^[l] Here again, Radday opens himself to charges that his work misses actual alternative explanations for his findings (Type II error).²²⁷

The Latter Radday. Linguistically, the latter Radday markedly improved upon the earlier Radday by rejecting his previous use of sentence lengths due to the critique of Buch,²²⁸ as well as rejecting his own prior use of word and parasyllabic interval lengths and phonemes based on his work on Zechariah.²²⁹ Thus, in the best tradition of exploratory empiricism, Radday increased

²²³ Radday, “Isaiah and the Computer,” 72.

²²⁴ Radday, “Isaiah and the Computer,” 73.

²²⁵ Radday followed Herdan’s insightful exploration of entropy (although it could be argued that syllable entropy might be one of the most incongruous places to start given that many other linguistic measures (such as lexemes, inflected words, clause groups and the like) display far higher numbers of discrete instances (tokens) in the biblical texts. See Radday and Wickman, *The Unity of Isaiah*, 117–28; Radday and Wickmann, “Unity of Zechariah,” 34–35.

²²⁶ Radday and Shore, *Genesis: An Authorship Study*.

²²⁷ Radday, “Two Computerized Statistical-Linguistic Tests,” 324.

²²⁸ Buch, “A Note on Sentence-Length as Random Variable,” 272–75; Radday et al., “The Book of Judges Examined,” 477–78.

²²⁹ Radday and Wickmann, “Unity of Zechariah,” 33–34.

the depth and breadth of his linguistic measures,²³⁰ and dropped those that demonstrated poor discrimination. But Radday was not yet done. It was precisely here – experimentally—where the latter Radday made his most important improvements compared to his prior praxis. Three examples make this transformation clear. *First*, the earlier Radday seemed initially unaware of the literature of experimental design, while the latter Radday demonstrated sensitivity to both sample size and block design of experiments.^{231[n]} *Second*, the earlier Radday failed to explore underlying causes of variation (genre, audience, etc.) other than authorship, while the latter Radday made strides to experimentally compare counts *within* literary genres.^{232[o]} *Third*, the earlier Radday did little more than assign any discovered significance to authorship alone. Nowhere was this more typified than in his 1975 *Zeitschrift für die Alttestamentliche Wissenschaft* article on the unity of Zechariah. In that work Radday and his collaborators divide the text into four sections (I-IX) with sections I-II (chapters 1-8) putatively authored by Zechariah and sections III-IV (chapter 9-14) disputed. When Radday uncovered findings that placed sections III or IV either intermediate or statistically indistinguishable from I-II (driven by the emphatic particle,²³³ certain parts of speech,²³⁴ and P. Guiraud’s vocabulary concentration²³⁵) these measures were dropped or assigned to other causes. Conversely, when linguistic measures in sections III or IV demonstrated distinctiveness from I-II, these measures are accepted as authorial.²³⁶ Thus, in his earlier work, Radday efficiently collected, in this case, only those

²³⁰ E.g. Radday added the definite article (Radday et al., “The Book of Judges Examined,” 479–80) as well as discourse features (albeit defined using traditional grammar); see Radday and Shore, *Genesis: An Authorship Study*, 30.

²³¹ For an example of Radday’s block design see Radday and Shore, *Genesis: An Authorship Study*, 21–23. For his sample design see Radday and Shore, *Genesis: An Authorship Study*, 24–27.

²³² Radday and Shore, *Genesis: An Authorship Study*, 19.

²³³ Radday and Wickmann, “Unity of Zechariah,” 36.

²³⁴ Radday and Wickmann, “Unity of Zechariah,” 38.

²³⁵ Radday and Wickmann, “Unity of Zechariah,” 41.

²³⁶ Radday senses this objection and states, “one may ask whether this may not be attributed to subject-matter or genre,” and then argues that genre is ill-defined in scholarship. Even if this were granted, it quite misses the point,

measures that supported his point and then calculated their joint likelihood.²³⁷ Seven years later, however, Radday explored Genesis using three causal groupings (the familiar JEP source/authorship framework) and two quasi-frameworks for what today we might label components of genre. The first of these components he labeled “Sorts of Discourse” or SDS. SDS includes three components: narrator, *dramatis personae*, and divine direct speech (NHD). For the second quasi-framework of genre, Radday appropriated the familiar literary divisions in the Genesis text, which he abbreviated as DIV. These Radday further subcategorized as *Urgeschichte* (chapters 1-11, DIV I), the Patriarchal Cycle (chapters 12-36, DIV II) and the Joseph Story (chapters 37-50, DIV III). Once Radday analyzed the Genesis text by these three frameworks, the Hebrew scholar could barely contain his surprise, “The chart reveals the startling fact that at least as much variation in language behavior is... [from sorts of discourse, SDS] as in Fig. 1 [the documentary hypothesis of authorship (DOC)].”²³⁸ This was the first discovery in the biblical texts that grouping the text by even a nontraditional literary definition of genre may constitute a more parsimonious explanation of the data than authorship. It opened the door wide for the possibility that genre, or some sociologically proximate transform of it, may explain a larger percentage of the summative variation seen in Radday’s measures, and indeed elsewhere. Inspection of his three central charts (DOC, SDS, and DIV),²³⁹ moreover,

because the genre employed was assuredly not ill-defined to the Hebrew prophet (Radday and Wickmann, “Unity of Zechariah,” 42).

²³⁷ Given that he has assumed rather than demonstrated the association between his selected measures and authorship, and systematically eliminated others, Radday’s final calculation violates the randomized assumption central to the calculation of any joint distribution. For other statistical problems just as severe, see the critique of Portnoy and Petersen, “Biblical Texts and Statistical Analysis,” 13–15.

²³⁸ Radday et al., “Genesis, Wellhausen and the Computer,” 470. Note: the square brackets indicate this author’s clarifications.

²³⁹ Radday et al., “Genesis, Wellhausen and the Computer,” 471–73.

demonstrates that, if anything, Radday was overly conservative in evaluating the extent to which the SDS grouping served as a superior discriminator to the DOC grouping.²⁴⁰

What then, shall we make of Radday? Although the earlier Radday was prone to statistical errors²⁴¹ and simplistic linguistic designs, the latter Radday became a broader linguist and a more careful experimentalist. In these things, Radday becomes an archetype of how an already capable biblical scholar can evolve into an equally capable computational experimentalist.

3.8.6 Anthony Kenny

Anthony Kenny's *A Stylometric Study of the New Testament* is a statistical work on the NT like virtually no other. Those who have read Kenny elsewhere will not be surprised to learn that his effort is an engagingly written study by a careful scholar in an effort to systematically analyze the text of the GNT. What is most notable about the work, however, is that Kenny charted an entirely new course.²⁴² He did not begin by selecting markers *ad hoc* and testing them with some measure of inference. *Rather, Kenny began by exploring all instances in an aggregate grammatical class.*^[p] His approach was therefore systematic and whole distribution-based. After having made all his arguments, Kenny aptly characterized his praxis as that of an aerial photographer.²⁴³ From that altitude, Kenny systematically worked through the various categories of grammar that he believed to be stylistically relevant. In so doing Kenny charted a broadly

²⁴⁰ Both visually, and as measured by the appropriate statistic (the F ratio of simple Euclidean distance from the group centroid), SDS more efficiently explains more of the variation seen in the Genesis text than authorship. Radday et al., "Genesis, Wellhausen and the Computer," 471–72.

²⁴¹ Specifically, per Portnoy and Petersen, inappropriate use of tests of simple univariate inference, failure to adjust the serial use of those tests for the multiplication of Type I error, failure to use multivariate inference, and the failure to use Bayesian technique where they are wholly appropriate (Portnoy and Petersen, "Biblical Texts and Statistical Analysis," 16–17).

²⁴² Kenny was no newcomer to computational stylistics, having previously published *The Computation of Style*, an elementary univariate work on computational stylistics (Kenny, *The Computation of Style*, passim).

²⁴³ Kenny, *A Stylometric Study of the New Testament*, 116–22.

descriptivist course, and made his conclusions cautiously. By systematically observing high-level linguistic differences between the texts of the GNT, Kenny joined a growing minority of researchers who concluded that the observed differences were suggestive of causes other than idiolectic variation.²⁴⁴ In the work as a whole Kenny exercised a rare combination of clarity, brevity, and accuracy in addressing his topic. If I may be permitted freedom to shift his aerial metaphor, Kenny correctly sensed that the forest of language distributions should be understood before its individual trees. In my view, Kenny's unique contribution lay precisely here.

Kenny's work, however, displays a number of deficits as well. This is hardly a work of rigorous inference. Most of his tables are expressed in percentages with no inferential apparatus in evidence, even when one easily could have been provided. Nor is this a work of modern linguistics. Kenny viewed the Greek of the New Testament primarily lexically, and through the parts of speech of traditional grammar.²⁴⁵ Kenny's exploration of syntagmatics is even more cursory, limited to sentence lengths and, quite unfortunately, to Morton's *positional stylometry*.²⁴⁶ In terms of his conclusions, Kenny's "aerial overview" demonstrated that in some respects the traditional Pauline corpus was homogeneous and in others quite distinct—a finding that left Kenny with no compelling reason "to reject the hypothesis that twelve of the Pauline Epistles are the work of a single, unusually versatile author."²⁴⁷

Finally, the only design insight we have not yet drawn from Kenny's work is a quite positive one—his work breathed a fundamental, distributionally representative *and cautious*

²⁴⁴ Especially noteworthy for our purposes is Kenny's observation of the high percentages of adjectives in the Pastoral Epistles (Kenny, *A Stylometric Study of the New Testament*, 92–94). This is a characteristic of the genres of *paraenesis* in general and *mandata principis* in particular.

²⁴⁵ Kenny, *A Stylometric Study of the New Testament*, 32–71.

²⁴⁶ Kenny, *A Stylometric Study of the New Testament*, 101–15.

²⁴⁷ Kenny, *A Stylometric Study of the New Testament*, 100.

empiricism. By so doing he allowed the “markers” to emerge up from the contours in the text itself. This must be emulated in our design.

3.8.7 Summary Design Implications

Many of the design choices in mid-twentieth-century stylometry reflected what in hindsight constituted poor rather than best practices. As we have seen, some were simple failures to do the requisite bibliographical work. Others were more systematic. These findings collectively suggest that the following practices should be incorporated in our design:

3.8.7.1 In Terms of Language Measures

1. When selecting linguistic features for supervised classification, for either idiolect or sociolect, develop a representative methodology.²⁴⁸ In lieu of this, use more measures and not less (pace Morton).²⁴⁹
2. Choose high-frequency rather than low-frequency forms for authorship attribution discrimination (pace words that only occur once).²⁵⁰
3. Avoid picking measures *ad hoc* (pace Morton).²⁵¹

3.8.7.2 In Terms of Experimental Design

1. Systematically eliminate other components that may explain that variation.²⁵²
2. Do not assign the cause of stylistic variation *ad hoc* (pace Morton).
3. Ensure sufficient sample sizes (pace Bee’s cumulative sum approach).²⁵³
4. Apply appropriate statistical methods (pace Bee’s use of normal statistics on skew distributions).²⁵⁴
5. Implement validation schema per Forbes’ response to Radday.²⁵⁵

²⁴⁸ See Drake’s first unanswered question. Drake, “Unanswered Questions,” 241–42.

²⁴⁹ See Michaelson and Morton, “The New Stylometry,” 101–02.

²⁵⁰ Morton’s use of infrequent linguistic measures is in odd contrast to his own dictum that “to make any judgment which will have practical applications and reliable results, we must look at events which happen fairly often.”

Morton and Winspear, *It’s Greek to the Computer*, 47.

²⁵¹ Hence Morton’s linguistically unfortunate selection of first and last words in a sentence (Morton, “Once. A Test of Authorship,” 5). In picking markers without justifying their neutrality to epistolary subgenre, this confounds any possible interpretation for authorship.

²⁵² Or in Drake’s words, “...before we can fully justify Radday’s conclusion, we must eliminate all other reasonable explanations for any established variation in style.” Drake, “Unanswered Questions,” 241.

²⁵³ Bee, “A Statistical Study of the Sinai Pericope,” 408; Bee, “The Use of Statistical Methods in Old Testament Studies,” 260.

²⁵⁴ While correctly asserting that F and t-test display “well-known robustness...towards departures from the normal distribution,” Bee seems to have nowhere tested his distributions for normality. Bee, “Statistical Methods in the Study of the Masoretic Text,” 615. Such tests would seem to be called for given that the coefficient of variation from virtually all tables, save Table 7 (pages 618–21), seems to be substantially greater than unity.

²⁵⁵ Forbes, “A Critique of Statistical Approaches to the Isaiah Authorship Problem,” 531.

3.8.7.3 In Terms of Conclusions

1. State conclusions conservatively, only after eliminating Type II error (pace Morton).
2. Control for statistical overstatement (pace Morton).

3.9 Milestone Four: The New Stylistics 1.0: Disentangling the Layers (Late 1980s)

Delcourt accurately signaled that a concern for fundamental issues characterized the period of the 1980s into the early 1990s.²⁵⁶ The issues were so fundamental, in fact, that it justifies viewing the period of the mid-1980s onward as the era of the *New Stylistics*. No less than six research programs that began in that period still continue actively today:

- The use of multivariate analysis to uncover language components or “layers”
- The use of more principled, linguistically-informed measures
- The use of automated feature set extraction (FSE) to objectively select linguistic measures
- The use of supervised mathematics in concert with FSE
- The use of different strategies of validation
- The emergent focus on latent structure analysis and causation

Said more turgidly, perhaps, in a discipline known for over-promising and under-delivering, the New Stylistics was a different kind of wineskin—it probed texts with new measures, selected those measures with new methods, assayed the contours of corpora with multivariate methods, and attempted to interpret those new findings with models that viewed language as intrinsically multilayered and multicausal.²⁵⁷

As with most Kuhnian paradigm shifts in the applied sciences—and the New Stylometry was decidedly that—most of the raw material for the revolution was already known in advance. All that was needed was a timely crisis and a new idea. That new idea will be treated first because history delivered it first. The “crisis” then will be reserved for Section 3.13.

²⁵⁶ “If expansion characterized the 70’s, a concern with fundamental issues characterized the 80’s.” Delcourt, “Stylometry,” 981.

²⁵⁷ This fundamental insight was not lost on Ledger who presciently stated, “Multivariate statistical analysis (MVA) has rarely been used in stylometry, although the latter is a discipline which seems to invite the application of such methods.” Ledger, “A New Approach to Stylometry,” 67.

3.9.1 The Early “Disentangling” (Component-Extraction) Studies

Prior to the mid-1980s the notion that language variation was multicausal was granted.²⁵⁸ Clear statements of this intuitive notion had already been issued by a wide mix of NT scholars, linguists, and stylometry researchers including but not limited to Nollen, Parry, Guthrie, Allen, Morton, and Mosteller.²⁵⁹ Intriguingly, however, although multiple causality was literarily and linguistically granted, it was all but experimentally ignored. This was a clear missed opportunity in stylistics given that more than enough evidence had accumulated by the mid-1980s to test this insight. This body of evidence included, for instance, Carroll’s seminal study on the analysis of prose (1960), Blackith’s analysis of Latin Elegiac verse (1963), Moerks’s description of style (1973), Brainerd’s distinction between novel and Romance (1973), Ule’s clustering of Elizabethan texts (1974), Boreland and Galloway’s attempt at author discrimination (1980), Faigley and Meyer’s important linguistically informed classification of text types (1983), Biber’s study of macroscopic text variation (1985) and Ledger’s proposal to use multivariate statistical analysis for stylometry (1985).²⁶⁰ Though the work of these scholars clearly lay outside the main channel of stylometric theory and praxis, it is they who first discovered that *layers of textual variation* could be extracted from texts. These scholars, from the periphery at least, sensitized their colleagues to the notion that stylistic variation was multicausal and that genre could

²⁵⁸ Although the *de facto* mainstream presumption however, was that authorship was the 900 pound gorilla in the room of language variation.

²⁵⁹ See especially, Nollen, “Qualities of Style as a Test of Authorship,” lxx; Parry, *The Pastoral Epistles with Introduction*, cxxvi; Guthrie, *The Pastoral Epistles and the Mind of Paul*, 27; Allen, “Methods of Author Identification,” 906; Morton, “The Authorship of Greek Prose,” 170; Mosteller, “The Writing Styles of the Authors of ‘the Federalist’ Papers,” 140.

²⁶⁰ For a description of the methodologies and their interpretation see Carroll, “Vectors of Prose Style,” 283–92; Blackith, “A Multivariate Analysis of Latin Elegiac Verse,” 200–04; Brainerd, “On the Distinction between a Novel and a Romance,” 260–69; Moerk, “An Objective, Statistical Description of Style,” 51–57; Brainerd, “On the Distinction between a Novel and a Romance,” 260–69; Ule, “Cluster Analysis,” 16–21; Boreland and Galloway, “Authorship, Discrimination, and Clustering,” 125–51; Faigley and Meyer, “Rhetorical Theory and Readers’ Classifications of Text Types,” 309–24; Biber, “Investigating Macroscopic Textual Variation,” 340–57; Ledger, “A New Approach to Stylometry,” 68–71.

confound the interpretation of authorship. What was unclear to most practicing stylometricians, however, was what multivariate analysis could possibly reveal about the *relationship* between genre and authorship. Or anything else for that matter. Enter J. F. Burrows.

3.9.2 The Pied Piper of the New Multivariate Stylistics: John F. Burrows

3.9.2.1 What He Did

In 1987, in both his book and his LLC article, Burrows outlined the major features of what would, discontinuously but emergently, coalesce into the multivariate arm of the New Stylistics.²⁶¹ Burrow's goal was simple; he sought to disentangle the layers of the text, motivated in part by Bakhtin's characterization of genre as a "stylistic medley." His strategy was to develop "more adequate methods....capable of comparing the languages of different genres, different historical periods, or different authors."²⁶² Secondly—and this constituted a sea change in stylistic studies—Burrows rejected the naive assumption that *ad hoc* markers "marked" whatever the researcher said they did. Instead, his strategy was to work from the phenomena of the text upward rather than from the researcher's assumptions downward. In his landmark article "Word Patterns and Story Shapes: The Statistical Analysis of Narrative Style," Burrows:²⁶³

1. Measured the top thirty inflected "word-types" (inflected words) by frequency.
2. Submitted these thirty inflected word-types to PCA (which Burrows called eigen-analysis).²⁶⁴ Recall that PCA extracts sequential components—usually weighed combinations of linguistic "features" such that the first component accounts for the most variation, and the second component is derived from the remaining variance and so forth. (See section 3.9.3 below.)
3. Plotted each text using the first two components of the PCA.

²⁶¹ Burrows, *Computation into Criticism*, passim; Burrows, "The Statistical Analysis of Narrative Style," passim.

²⁶² Burrows, "The Statistical Analysis of Narrative Style," 61.

²⁶³ Burrows, "The Statistical Analysis of Narrative Style," 62–70.

²⁶⁴ Neither in Burrows' article nor in his parallel monograph, *Computation into Criticism*, does he directly reveal the flavor of EMVA used, though it is revealed to be Principal Component Analysis (PCA) due to a label discovered on one of his exhibits (Burrows, *Computation into Criticism*, xv).

4. Inspected the groupings of the texts. In so doing, Burrows resolved the three text-types well known to readers of Jane Austen: dialogue, character narrative, and pure narrative.²⁶⁵
5. Used only the data characterized as pure narrative and performed the procedure over again, yielding (with one exception) a chronological sequence of Austen writings.²⁶⁶
6. Used the same process to demonstrate that when controlled for genre (Burrows used pure narrative) PCA successfully separated contemporaneous authors (James, Heyer, and Forster) from Austen.²⁶⁷

3.9.2.2 What Did Burrows Achieve?

First, Burrows' work should be recognized for what it was *not*. As we have already seen, Burrows was neither the first to call for a bottom-up analytical strategy in stylistics, nor the first to execute it. What he did do was to *popularize* four ground-clearing insights pioneered by his predecessors, assemble them all in one place, and explore their initial implications for sociocontextual layers of language.²⁶⁸ *First*, Burrows used linguistic measures representatively (by frequency order) and not selectively.^[a] Thus, his findings emerged upwards from language in use, rather than downward from disembodied theory. *Second*, by using multivariate analysis, he tested (rather than assumed) whether the language was multifactorial in nature and whether the factors so extracted were interpretable.^[b] *Third*, he drilled down within a given text-type, the genre of narrative, and executed the analysis yet again.^[c] Colloquially this can be understood as a sort of mathematical process of elimination. By eliminating variation caused by other genres, Burrows allowed the mathematics to do its inductive exploration unconfounded by variation caused by heterogeneity of genre. This allowed subsidiary underlying structural layers to

²⁶⁵ Burrows, "The Statistical Analysis of Narrative Style," 64.

²⁶⁶ Burrows, "The Statistical Analysis of Narrative Style," 64–66.

²⁶⁷ Burrows, "The Statistical Analysis of Narrative Style," 66–69

²⁶⁸ Although Burrows was not the first to apply decompositional mathematics to texts, he was the first to do so with broad linguistic measures derivative of linguistic findings (the layers of textual variation). Both Campbell and Love concur with our estimate that Burrows' work was pioneering in stylometry (Campbell et al., "The Provenance of *De Doctrina Christiana*," 106; Love, *Attributing Authorship: An Introduction*, 142).

emerge—and what emerged as the *second* derived component was *authorial variation*. This finding is central to our thesis and needs restating: Either *the major moment (eigen-systems derived component) of variation in mixed genre corpora tends (a) to be genre first, followed by authorship* or *(b) the combined biplot of the first two components tends to conformationally reflect genre rather than authorship*.²⁶⁹ Fourth, Burrows demonstrated that EMVA mathematics could separate and recover phenomenally what many have sensed epiphenomenally about texts since texts became texts. That is, they encode the things that literary theorists had been telling us were there all along (e.g. genre, date, and authorship). Fifth, Burrows provided experimental evidence that enabled informed scholars to finally reject the *de facto* fiction that style was monolithic in composition²⁷⁰ and idiolectic in source.²⁷¹

3.9.3 The Floodgates of Extractive Multivariate Analysis (EMVA) Open

After Burrows the trickle of EMVA studies became a flood. Before 1988 the EMVA computational stylistics literature totaled, by my count, a mere sixteen studies. By the fall of 2014 there were 146 such studies.²⁷² In fact, from the period of 1987 until about 2010,²⁷³ the rate of new eigen-systems studies doubled almost every seven years. By that same rough count, not counting reviews, responses, introductions and the like, from 1987 to early 2014, there were 130

²⁶⁹ In Hoover's words, "...genre differences are typically great enough to overwhelm the subtler differences between author's styles." Hoover, "Multivariate Analysis and the Study of Style Variation," 342.

²⁷⁰ Burrows' early discovery that even within a work an author does not display stylistic homogeneity was later supported by others. See Gurney and Gurney, "Subsets and Homogeneity," 139.

²⁷¹ Later Burrows, ever the English scholar, would write, "The most powerful differential I have encountered is the simple contrast between direct and indirect speech. When that is brought under control, either by an appropriate choice of texts or by the exclusion of the words that it most affects, differences of genre come into their own...In the Ocean of the language...are the Other Worlds and Other Seas of human individuality." Burrows, "An Ocean Where Each Kind," 319.

²⁷² This includes only those that for our purposes are stylometrically relevant: authorship, genre, dating, gender and other aspects of sociolect. See Table 2.2.

²⁷³ By 2010 the rate of production of eigen-systems studies slowed in deference to machine learning methods. See Section 3.14.

relevant eigen-systems-predicated stylometric studies published, including 66 PCA studies,²⁷⁴ 12 FA studies, 70 MDA studies, and 10 CA or MCA studies.²⁷⁵ In an attempt to summarize this literature there are at least two significant challenges. First, executing a comprehensive summary would require perhaps an additional 100-page review! Second, these EMVA studies were executed using different linguistic measures, on different languages, for different purposes, using different EMVA approaches and executed with disparate designs. Therefore, even though 146 relevant studies have been executed, we must be cautious about their generality. What does, for instance, Mealand's correspondence analysis (CA) of the Pauline Canon mean given his mix of linguistic measures?²⁷⁶ Would his findings have been similar at each distinct level of linguistic rank? Or for all systems of the Koine? Or for all selection expressions at a given level of delicacy in the systems network of the GNT? What was stated in section 2.3 bears underscoring here: "linguistically, no studies have attempted to address a relevant problem in New Testament studies *by using a comprehensive set of linguistically-derived measures for either the syntagmatic structures, paradigmatic systems, multiple strata or multiple metafunctions of the GNT.*"²⁷⁷ The same can be said of studies outside of the GNT. Seen in this way these 146 widely

²⁷⁴ See especially the early work immediately following Burrows; Brunet, "*Le vocabulaire religieux*," 161–65; Loehlin, "Component Analysis Versus Common Factor Analysis," 29–31; Ledger and Merriam, "Shakespeare, Fletcher, and the Two Noble Kinsmen," 239–41; Holmes and Forsyth, "The Federalist Revisited," 117–21; Ledger, "An Exploration of Differences in the Pauline Epistles," 87–95; Baayen et al., "Outside the Cave of Shadows: Using Syntactic Annotation," 122–28; Mealand, "The Extent of the Pauline Corpus," 73–80; Temple, "A Multivariate Synthesis," 68–74. For the most accessible mathematical introduction for PCA applied to stylometry see Binongo and Smith, "The Application of Principal Component Analysis," 447–60.

²⁷⁵ The number sums to more than 130 because some studies use multiple EMVA methods.

²⁷⁶ Mealand, "The Extent of the Pauline Corpus," 74–86.

²⁷⁷ Mealand may have been the closest to having done this, exploring as he did separate correspondence analyses of connectives: particles, conjunctions, and prepositions in the GNT (Mealand, "Style, Genre, and Authorship in Acts," 494–99). While this demonstrates Mealand's clear understanding that multivariate analysis must be executed at the level of coherent linguistic species, neither can such work be called comprehensive. In the work in Chapter Five, for instance, HLLA models have been built upon all 51,599 species in the rank structure of the GNT (species actually found at least once in the GNT). Assuming that one-tenth of those are found in high enough numbers to be statistically tested, this means that there are on the order of over 5,000 rank measures alone that can be assayed in the GNT.

dispersed studies are relatively early phase attempts to disentangle idiolect from sociolect. With this caution duly registered, four summary statements—themes if you will—can be issued from these EMVA findings.

3.9.4 A Summary of the EMVA Findings Executed to Date

First, more frequently than not the primary component (dimension) in unsupervised exploratory decompositional data analysis is *genre*.²⁷⁸ The importance of this finding should not be ignored. Recall from our first chapter that if this is found to be true of the GNT as well, to the extent that prior studies have failed to test for genre as a competing theory or failed to remove genre as a covariate, those studies have almost certainly confounded genre with authorship. *Second*, the sociolectic story doesn't stop there. EMVA has extracted other sociolectic factors from these texts other than literary genre,²⁷⁹ including gender,²⁸⁰ date (stylochronometry),²⁸¹ and other factors endemic to the literature examined.²⁸² *Third*, computational stylistics began to use far more markers than ever before.²⁸³ *Fourth*, generally though unevenly, computational

²⁷⁸ For example, Burrows, "The Statistical Analysis of Narrative Style," 64; Forsyth et al., "Investigating the Authenticity of the Consolatio," 383; Juola and Baayen, "Authorship Identification by Cross-Entropy," 63; Baayen et al., "An Experiment in Authorship Attribution," no pages, section 3. Cf. also Burrows, "The Interpretative Nexus between Analysis and Information," 92–102. See also Craig, who while cautious in assigning generic differences notes that the first component in his PCA analysis distinguished most clearly "between the history plays and the others." Craig, "Stylistic Analysis and Authorship Studies," 274.

²⁷⁹ Our meta-analysis details thirty-three studies related to the extraction of genre from texts. See especially the inaugural work of Burrows (Burrows, "The Statistical Analysis of Narrative Style," 64) and the following studies: Biber, *Variation Across Speech and Writing*, 170–98; Burrows, "The Interpretative Nexus between Analysis and Information," 99–103; Mealand, "Measuring Genre Differences in Mark," 229–42; Mealand, "Style, Genre, and Authorship in Acts," 489–99.

²⁸⁰ Burrows, "An Ocean Where Each Kind," 311; Thomson and Murachver, "Predicting Gender," 198–204; Kucukyilmaz et al., "Chat Mining for Gender Prediction," 280–82.

²⁸¹ See especially Temple, "A Multivariate Synthesis," 69–74; Forsyth et al., "Investigating the Authenticity of the Consolatio," 379–96; Can and Patton, "Change of Writing Style with Time," 66–77.

²⁸² Arguably Lana's dimensions using correspondence analysis separate *topics* (Lana, "Xenophon's Athenaion Politeia," 21).

²⁸³ Burrows and others began to use increasingly larger percentages of the tokens in their corpora. In his 1989 CH article he used fully forty percent of the word-token in Austen's novels; Burrows, "An Ocean Where Each Kind," 309.

stylistics became more cautious in its conclusions—a reflection, in my view, of the growing recognition of the nuanced complexity of the linguistic quarry being pursued.²⁸⁴

3.9.5 Design Implications

Pragmatically, given the theory defined above, the following implications should be folded into our experimental design:

1. All linguistic forms (tokens) to date encode some sort of information; therefore all “levels” of all the modes of language (stratal, paradigmatic, syntagmatic, and metafunctional) should be systematically assayed.^{285[d]}
2. All three central activities of computational stylistics should be executed:²⁸⁶
 - a. *Analysis*: Decompositional Analytics via EMVA to understand the “layers”^[e]
 - b. *Synthesis*: Supervised approaches (which we will describe shortly) should be used in concert with EMVA to solve classification problems (e.g. authorship)^[f]
 - c. *Causation*: Multivariate causal mathematics should be executed both in terms of EMVA studies (especially Factor Analysis) and more advanced methods of causal modeling (e.g. LISREL, SEM and the like)^[g]

3.10 Period Five: The Interregnum That Led to the Wild West (Mid-1980s–1990s)

Section 3.7 demonstrated that the careful experimental design contributions of Mosteller and Wallace in the 1960s were universally lauded, but seldom imitated. Similarly, many researchers by the end of the 1980s recognized that EMVA work by researchers such as Burrows successfully mapped to sociocultural layers in language. The problem? Unlike the mathematics of Mosteller and Wallace, this insight complicated rather than simplified the then marquee objective of stylometry: authorship attribution. Worse yet, it raised the embarrassing specter that some, if not many, of the vaunted statistical differences developed to date by univariate methods

²⁸⁴ Typical of this far more conservative approach to conclusions was Frautschi’s work on the prose tales of Perrault, “While the results of the present study provide no conclusive evidence regarding the authorship issue, the identification of two alternating focal styles offers a direction for further stylistic analysis.” Frautschi, “Focal Style and the Problem of Attribution,” 213.

²⁸⁵ Burrows, “The Statistical Analysis of Narrative Style,” 62.

²⁸⁶ Burrows terms the first two tasks the descriptive and the predictive (Burrows, “The Interpretative Nexus between Analysis and Information,” 91–99).

may have been from causes *other* than authorship.²⁸⁷ Whether for this reason or in spite of it, the emergence of the multivariate arm of the New Stylistics marked the beginning of a clear interregal period. There was an emerging recognition that the old king—the straight line linkage between significant stylistic variation and authorship—had died, but the new crown prince of multivariate stylistics had yet to experience his investiture. More tellingly, who would this new crown prince be anyway? And how would the discipline tease apart these confounding effects to get back to work on authorship attribution? These questions—to say nothing of the intimidating new methods involved—complicated and delayed the quick acceptance of Burrows’ approach among casual and even inveterate practitioners of computational stylistics. Consequently, in the beginning of this period in particular, much computational stylistics work by sheer momentum still:

- Used language measures defined only by traditional grammar and syntax
- Used questionable linguistic measures (sentence lengths and the like)
- Selected only a handful of measures, opening themselves to the charge of what later would—accurately—be termed “cherry-picking”²⁸⁸
- Persisted in using univariate inference rather than multivariate analysis
- Showed little interest in disentangling sociocultural or sociolinguistic causes of variance from idiolectal variation

In terms of our more focused interests on sociolect and its bearing on the GNT, three themes characterized the mid-1980s to the mid-1990s: (1) the final charge of serial studies in univariate analysis, (2) the period of the “Wild West” in the early multivariate analysis of texts and (3) the emergence of the New Stylistics in our own discipline of NT studies.

²⁸⁷ Mealand raised this central point in his 1988 JSNTS review article (Mealand, “Computers in New Testament Research,” 108). See also his comments on “making allowance for genre” in Mealand, “Positional Stylometry Reassessed,” 286.

²⁸⁸ See especially Foster, “The Claremont Shakespeare Authorship Clinic,” 500; Craig, “Stylistic Analysis and Authorship Studies,” 286.

3.10.1 The Final Charge of Univariate Analysis

We face a quandary in reviewing the first decade of this period and its central figures, Thomas Merriam and M.W.A. Smith. On the one hand, this period and these two researchers constituted one of the main channels of stylometric research from the mid-1980s to the mid-1990s—and both scholars were capable and assiduous researchers. On the other hand, because the methods they initially adopted (and would grow beyond) occupied most of their efforts during this period—and have been largely abandoned—we must restrict ourselves to making only four comments.

First, though Merriam and Smith could not have known it at the time, their efforts would comprise the final mainstream attempt to triangulate computational stylistics to the theories and methods of A.Q. Morton. In so doing both scholars approached their task with different starting points. Merriam, fundamentally, sought to test Morton's methods and improve upon him. He initially accepted Morton's reasonable but oversimplified dictum that "stylometry is based on the hypothesis that habit variation within the same author and same genre cannot be distinguished from random variation."²⁸⁹ M.W.A. Smith did not initially disagree and even used Morton's method himself.²⁹⁰ Yet, Smith, ever the careful experimentalist, was quicker to conclude that Morton's methods were simply not up to the task. These different propensities, Merriam seeing Morton through his possibilities, and Smith seeing Morton through his problems, needed only a spark to become a full-fledged academic conflagration. *Second*, that spark came when Nigel

²⁸⁹ Merriam, "An Experiment with the Federalist Papers," 252. In fact, different authors evidence widely different quantities of intra-authorial variation, even within the same genre. Said colloquially, some authors write quite woodenly, others vary their style widely even within a single genre. An example of this can be seen in Smith and Rickard's later work in which advanced mathematics achieved the maximum separation between the known works of Thomas Jefferson and Thomas Paine. When applied to the Declaration of Independence it demonstrated, erroneously, that Paine, not Jefferson penned the Declaration! Smith and Rickards, "The Authorship of the American Declaration of Independence," 25. The conclusion? Creative writers *can* adapt their writing style in a way that obliterates the narrow, intra-authorial assumptions of Morton.

²⁹⁰ Smith, "The Authorship of a Lover's Complaint," 23.

Hawkes and Merriam separately used Morton's method to produce results directly contradictory to the majority opinion (and Smith's own views) regarding the Shakespearean apocrypha.²⁹¹ Smith for his part, explored the reliability of Morton's claims and concluded that inadequate experimental and statistical errors dogged both Morton's work and his conclusions.²⁹² Accordingly, rather than attempting to revitalize Morton's method, Smith chose to develop his own. Intriguingly his "new" approach, however, was not fundamentally different quantitatively—he used essentially the same chi-squared based approach as Morton—but chose to focus on different markers—the first word of speeches.²⁹³ *Third*, despite their different perspectives on Morton's methods, the contest resolved itself far more amicably than where it began. The reason is not difficult to discover. Although Morton's framework initially limited both researchers, by their sheer dogged empiricism they both converged upon a consensus of sorts, if not on Shakespeare, at least on Morton's limitations and his unqualified claims. In Smith's accurate assessment, Morton's method lacked the "power to discriminate" between some authors, but it nonetheless demonstrated "some element of discrimination."²⁹⁴ *Fourth*, Merriam and Smith's differing personalities and perspectives drove and influenced, as it always will, the exercise of empiricism. Smith stressed experimental design and statistical rigor, and his approach and dialectic was the more polemic and deconstructive of the two.²⁹⁵ Merriam, though he began

²⁹¹ Specifically, Merriam used thirty-two lexical tokens or ratios of tokens and concluded that over ninety percent of Sir Thomas More was from the mind (though not the hand) of Shakespeare (Merriam, "The Authorship of Sir Thomas More," 3).

²⁹² Smith, "An Investigation of Morton's Method," 3–10.

²⁹³ Smith's linguistic choice here is problematic at best. If one is hoping to propose a purely idiolectic Shakespeare, then the first words of speeches—a measure thoroughly influenced by the subject matter and audience of the speech—is an odd choice; Smith, "The Authorship of Acts I and II of 'Pericles'," 24. This concern is magnified given that some of the plays selected for authorial analysis are from different subgenres; Smith, "The Authorship of Acts I and II of 'Pericles'," 32.

²⁹⁴ Smith, "Shakespeare, Stylometry and 'Sir Thomas More'," 437. Holmes, correctly, concurred with Smith's negative assessment of Morton's statistical design choices. Cf. Holmes, "A Stylometric Analysis of Mormon Scripture," 92.

²⁹⁵ Smith was not above sidestepping the force of Merriam's point and inferring that the qualities of his deductions were suspect; Smith, "Merriam's Applications of Morton's Method," 60.

more tied to the oversimplifications of Morton, arguably evolved into the more creative experimentalist. Merriam, in particular, with the exception of some unconvincing exercises in letter frequency analysis and cumulative sums,²⁹⁶ emerged into a broad and productive period of the exploratory multivariate analysis of texts.²⁹⁷ Smith was early to embrace multivariate analysis as well,²⁹⁸ although his other choices—explorations of low-frequency word comparisons (following Slater’s work)²⁹⁹ first words of speeches and pairs of consecutive words³⁰⁰—belated his entry into the more fruitful multivariate work he would later publish with Jose Binongo.³⁰¹

3.10.2 The Period of the Wild West in Multivariate Stylistics

The paroxysms of change after Burrows, of course, went far beyond Merriam and Smith. While Merriam and Smith moved steadily, albeit unevenly, toward methods of multivariate analysis, much of the rest was a Kuhnian landscape—a veritable Wild West that would lead inexorably to a crisis in the discipline (Section 3.11). Positively, however, the emergence of EMVA, as well as other influences,³⁰² resulted in researchers using an increased number of

²⁹⁶ Merriam, in a textbook example of overfitting, found two letters that separated Marlowe and Shakespeare (Merriam, “Letter Frequency as a Discriminator of Authors,” 469; Merriam, “Heterogeneous Authorship in Early Shakespeare,” 18–21). He also explored cumulative sum charts without assaying alternative causes of their slope (Merriam, “Edward III,” 162–77; Merriam, “Marlowe and Nashe,” 427).

²⁹⁷ Merriam used principle components analysis upon Shakespeare (Merriam, “Invalidation Reappraised,” 420–29; Merriam, “Tamburlaine Stalks in ‘Henry VI,’” 270–77; Merriam, “Marlowe’s Hand in Edward III,” 19–21; Merriam, “Heterogeneous Authorship in Early Shakespeare,” 19–24; Merriam, “Edward III,” 164–72; Merriam, “An Unwarranted Assumption,” 439–40); as well as neural networks (Matthews and Merriam, “Neural Computation in Stylometry I,” *passim*; Merriam and Matthews, “Neural Computation in Stylometry II,” *passim*; Merriam, “Heterogeneous Authorship in Early Shakespeare,” 19; Merriam, “Edward III,” 182–83).

²⁹⁸ Smith, “The Authorship of ‘The Revenger’s Tragedy,’” *passim*; Smith, “The Problem of Acts I–II of Pericles,” *passim*; Smith, “Edmund Ironside,” *passim*.

²⁹⁹ Smith’s forays into work-links included Smith, “Word-Links and Shakespearian Authorship and Chronology,” *passim*, and Smith, “Word-Links and the Authorship of Edmund Ironside,” *passim*.

³⁰⁰ Smith, “A Procedure to Determine Authorship Using Pairs of Consecutive Words,” 114.

³⁰¹ See especially Binongo and Smith, “The Application of Principal Component Analysis,” *passim*; Binongo and Smith, “A Bridge between Statistics and Literature,” *passim*; Binongo and Smith, “Project Jacobean: A Mathematical Exploration of a Literary Era,” *passim*.

³⁰² Among these influences are (1) a more linguistic and literary view of language and subsequent tagging of texts with such content (e.g. Baird, “Content Analysis and the Computer,” 260–65), (2) low-cost access to powerful statistical packages (e.g. SAS, SPSS, BDMP) and (3) the emergence of professional organizations that established standards and provided publishing platform for results (e.g. the Association for Machine Translation and Computational Linguistics founded in 1962 and renamed to the Association for Computational Linguistics in 1968, as well as the founding of the Association for Literary and Linguistic Computing in 1973).

measures with increasing numbers of multivariate techniques and with an increased openness to sociocultural sources of variation. Less positively, however, this period was characterized by problems that looked very much like the same square of opposition we saw almost a century prior (see section 3.3) but this time with the breadth of language measures on one axis and the use of multiple EMVA measures on the other. We will treat three of the four sides of this square next.

3.10.2.1 Inadequate Linguistic Measures but Analyzed with Multiple Multivariate Methods

We start this section with the 1992 article by D.I. Holmes, “The Stylometric Analysis of Mormon Scripture.”³⁰³ This otherwise excellent work³⁰⁴ displays sensitivity to both the then current methodologies and word distributions. Anomalously, however, Holmes ended up executing his work only upon the five variables^[a] largely derived from the Sichel distribution,^[b] and this derivation was predicated upon only one part of speech (nouns). Here too we find Mealand’s LLC article comparing Acts, the Septuagint, and Hellenistic historians, which are detailed in Section 3.10.3.1. Located here as well is Greenwood’s thin set of the ten most frequent words in the traditional Pauline texts. Greenwood’s triad of articles will be discussed in Section 3.10.3.2. Mention should be made of Larsen’s effort to discriminate between putative authors in the Book of Mormon.³⁰⁵ Although his work technically precedes our period by five years or so (1980), it displayed important and careful concern for experimental design, on the one hand, and yet was linguistically sparse, accessing only lexical instances, and then only nine words chosen by MANOVA. Next, Tweedie, Holmes, and Corn similarly used only lexical

³⁰³ Holmes, “A Stylometric Analysis of Mormon Scripture,” 92–96.

³⁰⁴ Holmes’ work moved the bar forward by analyzing metadata and summary parameters of twenty-four noun distributions. Those parameters included Honore’s R, V2/ V, Yule’s K and 10,000-equivalent values of a and 0 in the Sichel model.

³⁰⁵ Larsen et al., “Who Wrote the Book of Mormon?,” 235.

frequencies in their study of the authorship of *De Doctrina Christiana*,³⁰⁶ but analyzed with multiple MVA procedures (PCA and cluster analytics). This same asymmetry between measures and methods was also in evidence with Jin and Murakami regarding commas,³⁰⁷ Ledger and Merriam's study with letter frequencies,³⁰⁸ and other studies including those of Forsyth, Holmes, and Stewart.³⁰⁹

3.10.2.2 More Adequate Linguistics Measures but Analyzed with Only One Multivariate Method

We look next at studies which used a more expansive set of discrete markers, but analyzed with only one method of multivariate analysis.^[c] This would include Merriam's authorship study of Edward III, a work which used only the first component of a PCA analysis supported by what I have argued elsewhere is, at best, a dubious technique: QSUM charts.³¹⁰ Dixon and Mannion, in a better-designed study, used thirty-nine separate measures, one that spanned, roughly, the rank scale of nineteenth-century English literature.³¹¹ Their analysis, however, was limited to cluster analysis techniques.³¹² Lastly, I highlight the article of Cooper and Pearsall.³¹³ It could be well argued that these studies display both inadequate linguistic measures and multivariate methods, given that their choice of measures largely included scalars (e.g. syllable counts)^[d] with their attendant poor discrimination as well as idiosyncratic measures

³⁰⁶ Tweedie et al., "The Provenance of *De Doctrina Christiana*," 78–80.

³⁰⁷ Jin and Murakami, "Authors' Characteristic Writing Styles," 63–65.

³⁰⁸ Ledger and Merriam, "Shakespeare, Fletcher, and the Two Noble Kinsmen," 237–38. Ledger's later LLC article displays the same limited letter-based measures but supplemented with the Type/Token ratio. Ledger, "An Exploration of Differences in the Pauline Epistles," 86.

³⁰⁹ Forsyth et al., "Investigating the Authenticity of the *Consolatio*," 379–94; Holmes et al., "Stylometry and the American Civil War," 406–09; Stewart, "Charles Brockden Brown," 131.

³¹⁰ Merriam, "Edward III," 172.

³¹¹ Dixon and Mannion, "Goldsmith and the *British Magazine*," 39–40.

³¹² Dixon and Mannion, "Goldsmith and the *British Magazine*," 40–44.

³¹³ Cooper and Pearsall, "The *Gawain Poems*: A Statistical Approach," 382–83.

(alliteration). In addition their method of MVA extraction lies deeply shrouded in mystery.³¹⁴

These examples are represented by dozens of other studies executed using cluster analysis.³¹⁵

3.10.2.3 More Adequate Linguistic Measures Analyzed with Multiple Methods of MVA

Lastly, our journey around the square of opposition now brings us to our preferred location, studies that display both multiple linguistic measures and multiple intersecting MVA analytics. This preferred location includes Stratil and Oakley's article on the plays of Tirso Molina. These scholars chose many measures to be sure, but not through any recognizable strategy of feature set selection.³¹⁶ Here, too, we find the excellently designed work by Maurizio Lana³¹⁷ who explored the authorship of works by Xenophon, Thucydides, and Aristotle analyzed via correspondence analysis (CA). Empirically he discovered the expected result, that CA yielded more interpretable clusters of texts for those runs that incorporated multiple rather than single parts of speech. Somewhat idiosyncratically, however, he first selected 40 words, apart from any clear rationale, but redeemed that effort by executing the CA using the entire lexical stock of 30,745 words,^[e] and then reducing it by eliminating *hapax legomena*.³¹⁸ Located here as well are Gurney and Gurney's articles, efforts which deploy cluster analysis, PCA and MDA

³¹⁴ Cooper and Pearsall merely report quite incredibly that, "The nature of this mathematical procedure need not detain us here." Cooper and Pearsall, "The Gawain Poems: A Statistical Approach," 382. We assume it to be some flavor of discriminant analysis.

³¹⁵ We omit cluster analytics from consideration in this review for two reasons. First, cluster analytics, whether in its hierarchical agglomerative or iterative K-means incarnations, are not extractive techniques. Hence, they do not enable us to disentangle components in the data. Second, twenty years of work with cluster analysis has convinced me that the technique is very frangible, meaning that the cluster results differ greatly based on the algorithm, aggregation method and distance measure employed.

³¹⁶ Stratil and Oakley, "A Disputed Authorship Study of Two Plays," 153–59. This article is a curious patchwork that employed largely antiquated sentence and word length measures as well as a nonsystematic approach to word content (read functional) analysis. Their primary technique was cluster analysis with PCA deployed as a control.

³¹⁷ Lana, "Xenophon's Athenaion Politeia," *passim*.

³¹⁸ Lana, "Xenophon's Athenaion Politeia," 18–24.

against multiple linguistic measures.³¹⁹ While few followed this pair in some of their choices (such as sets of unique words or lemmas in successive sequences of only 100 words of text)³²⁰ their work did demonstrate the discriminatory power of a set of at least 25-30 high-frequency function words.³²¹ More importantly, as was previously mentioned, Gurney and Gurney were two of the first researchers to analyze an entire set of tokens in texts.

3.10.3 The Emergence of the New Stylistics 1.0 in the NT

The flurry of this early EMVA work did not leave the world of NT studies untouched. Despite the relative paucity of EMVA analytics applied to the GNT (by our count only sixteen such studies have been executed to date) several of these studies are quite important from the perspective of our design in Chapter Four. I will organize this section of the review by researcher.

3.10.3.1 David L. Mealand and the Greek of the New Testament

There is much that can be said in praise of Mealand's work. First, he used multivariate techniques, and often several flavors of them, to provide different vantage points into the GNT.³²² Second, Mealand's work always evidenced a careful concern for experimental design, such as testing for requisite statistical assumptions (multivariate normality and the like).^{323[f]} Third, from his earliest papers Mealand was fully sensitive to extra-authorial sources of linguistic

³¹⁹ These measures included very creative combinations of lemma and word frequencies, word and sentence lengths, and punctuation just to name just a few. See especially Gurney and Gurney, "Authorship Attribution of the *Scriptores Historiae Augustae*," 124.

³²⁰ Gurney and Gurney, "Authorship Attribution of the *Scriptores Historiae Augustae*," 123.

³²¹ Gurney and Gurney, "Authorship Attribution of the *Scriptores Historiae Augustae*," 129.

³²² In his earlier work Mealand first used cluster analysis and discriminant analysis; Mealand, "The Extent of the Pauline Corpus," 71–79. He later added correspondence analysis and general linear models (GLM) Mealand, "Measuring Genre Differences in Mark," 231–35; Mealand, "Style, Genre, and Authorship in Acts," 491–98; Mealand, "Is there Stylometric Evidence for Q?," 490–94. Intriguingly, Mealand himself recognized the limitations of using discriminant analysis. Cf. Mealand, "The Extent of the Pauline Corpus," 64.

³²³ See for instance, Mealand, "Style, Genre, and Authorship in Acts," 488–89.

variation.^{324[gl]} Fourth, and refreshingly, from the outset Mealand expressed his observations cautiously.

Even Mealand, however, can be challenged at several points regarding his work. Notably, especially in his earlier studies Mealand used very thin sets of linguistic data (e.g. only four conjunctions analyzed only via chi-square in one study, and a discriminant analysis built only upon six linguistic measures in another).^{325[h]} Moreover, in his later papers, although the number of his measures grew larger, his selection of markers seemed informed by the somewhat idiosyncratic measures of Ledger (letter frequencies)³²⁶ and was further compromised by limiting his linguistic quarry to traditional grammar.^{327[i]} Nor did Mealand attempt to down-select linguistically representative markers using feature set selection criteria.^{328[j]} In my view, some of his “surprise”³²⁹ findings were because his linguistic measures were not drawn representatively from the paradigmatic and syntagmatic axes of language. In short, Mealand’s idiosyncratic approach yielded a kind of linguistic cocktail. This had two untoward effects. First, it washed out the salutary differences that would have been observed had his design studied the systems and

³²⁴ Per Mealand’s comments, “Perhaps the stylistic differences could be explained by differences of situation, genre, mood or whatever” (Mealand, “Positional Stylometry Reassessed,” 285), and “It is important...that we should not leap from the detection of groupings to immediate conclusions about authorship” (Mealand, “The Extent of the Pauline Corpus,” 76). Despite this, however, Mealand, like other researchers, tended to inexorably tilt toward interpreting distances between text groups in terms of authorship in his conclusions. Moreover, Mealand’s view of genre seems, at least, to be drawn from literary theory for its main genres, followed by form critical categories for its subgenres; Mealand, “Measuring Genre Differences in Mark,” 227–28.

³²⁵ For his design see Mealand, “Positional Stylometry Reassessed,” 270–79. For the aforementioned discriminant analysis see Mealand, “The Extent of the Pauline Corpus,” 81.

³²⁶ Mealand, “The Extent of the Pauline Corpus,” 66.

³²⁷ Mealand, “Style, Genre, and Authorship in Acts,” 494–99.

³²⁸ This is not to say that Mealand failed to down-select markers based on statistical criteria such as that available in commercial implementations of stepwise discriminant analysis. These methods, however, only delete non-contributing markers via entry and removal criteria. They do not tell one whether the starting set of variables is representative of the linguistic potential to be had at any given layer of the language being evaluated. Mitigating this somewhat Mealand used thirty-two total measures and although he does not present many CA visualizations, they triangulate modestly well those we present in chapter eight (Mealand, “Style, Genre, and Authorship in Acts,” 492.

³²⁹ Mealand, “The Extent of the Pauline Corpus,” 79).

structures of language separately.³³⁰ Secondly, his *a priori* decision to excise certain sections from the text (quotations) complicates a straightforward assessment of the texts as received.³³¹

3.10.3.2 Kenneth Neumann and the Pauline Epistles

In 1988 the SBL dissertation series published Kenneth J. Neumann's Th.D. thesis entitled, *The Authenticity of the Pauline Epistles in the Light of Stylostatistical Analysis*. Neumann's work was, in many ways, a significant step forward. Six salutary factors merit Neumann's position as a bridge scholar to the new computational stylistics. First, Neumann provided an appropriately extensive (ninety-two page) review of past statistical studies. This distinguishes Neumann's effort from almost all prior studies whose literature review typically failed to cross disciplinary boundaries. Moreover, he organized that review by linguistic category. Such an organization suggests sensitivity to the possibility that language may disproportionately encode different kinds of information via different aspects of language—specifically via its systems, structure and strata. Second, he carefully collected a comparatively wide variety of language measures, fully 617 of them, which he termed *indices*.^{332[k]} Third, he demonstrated a deep sensitivity to experimental design in both his sampling approach and his inclusion of contemporary extrabiblical texts in the analysis.^{333[l]} Fourth, he recognized that genres and subgenres of literature exist with varying proximity to the Pauline epistles.^{334[m]} Fifth,

³³⁰ Mealand would later attempt to sort out the components of his cocktail, at least in terms of the categories of traditional grammar; Mealand, "Measuring Genre Differences in Mark," 495–99. It should be noted that even Dr. Mealand's latest work (2012) is cast solely in terms of traditional grammar; Mealand, "Hellenistic Greek and the New Testament," 324.

³³¹ Mealand extracts, for instance, quotations from the GNT texts, and this compromises interpretation of genres that are defined by their presence (Mealand, "The Extent of the Pauline Corpus," 66). Despite these limitations, I have great personal affinity for Mealand's work. My own exploratory multivariate data reduction work on the GNT was outlined in six unpublished papers presented from 2006–2009 at SBL. Similar to Mealand's later work, my work began with correspondence analysis. I chose, however, to down-select linguistic measures by AAVASR and test linguistic measures separately. See Chapter Eight for a fuller description.

³³² Neumann, *The Authenticity of the Pauline Epistles*, 118–24, 73.

³³³ Neumann, *The Authenticity of the Pauline Epistles*, 131–40.

³³⁴ Neumann, *The Authenticity of the Pauline Epistles*, 125–31.

he used a multivariate technique, discriminant analysis, with the appropriate but standard set of associated inferential measures associated with that choice. Sixth, his initial results indicated that his reduced variables “really separate types of writing style more than authors.”³³⁵ This coheres with Burrows’ original findings. For purposes of this work it is critical to note that *while genre was not his primary focus Neumann was one of the first to discover/recognize genre predominance in eigen-systems-based multivariate analysis of the GNT*. These implications are not lost on Neumann as he concludes “Genre analysis, rather than authorship, could be attempted. *It could even be the most helpful direction for research on the disputed letters.*”³³⁶

Despite the significant forward progress made by Neumann, my critique overlaps to a certain extent with that of Mealand. Three concerns in particular are noteworthy. First, Neumann’s language measures (indices) are derived entirely from the wells of traditional grammar.^{337[n]} This might have been anticipated given that no work on theoretical linguistics, the functional schools, or even transformational grammar is listed in his otherwise extensive bibliography. Second, per Mealand, Neumann assays his many measures simultaneously, pouring them together into a kind of “linguistic cocktail.” The only way this will not disproportionately tilt his results toward one causal notion or another is if his selection of measures are somehow serendipitously balanced. But what would guarantee this serendipity? Third, unlike Burrows and Mealand, Neumann rests his case only upon one multivariate technique, discriminant analysis.^[o] This choice evinces several problems: (a) Discriminant analysis was hardly the obvious choice even in Neumann’s day. Given that the raw data in textual analysis is frequencies, a nonparametric method of multivariate analysis such as

³³⁵ Neumann, *The Authenticity of the Pauline Epistles*, 184.

³³⁶ Italics mine. See Neumann, *The Authenticity of the Pauline Epistles*, 228.

³³⁷ Neumann, *The Authenticity of the Pauline Epistles*, 115–18.

correspondence analysis would have been the expected choice. (b) Neumann does not explore the extent to which his down-selection of measures prejudices the outcome. How, we must ask, has he guarded against the very real possibility that the resulting subset of variables, selected by the statistical method of F-ratio analysis, does not disproportionately preserve or destroy measures related to some underlying cause (authorship, genre, subgenre, and the like)?³³⁸ (c) The most troubling facet of Neumann's experimental design is that based on the description given, I still cannot tell whether Neumann has, quite accidentally, stacked the deck.³³⁹[p] That is, discriminant analysis, unlike other unsupervised multivariate data reduction (MDR) techniques, *requires* the researcher to *provide* the grouping categories. That is what makes MDA a *supervised* method. What I cannot tell from his description, despite repeated readings, is whether Neumann placed ancient texts together in modern text groupings before submitting them to discriminant analysis. Specifically, did Neumann set up a seven-group discriminant analysis (one group for each set of texts determined *a priori*) or did he execute an approximately thirty-plus group discriminant analysis (one group for each sample)? If the former, then in effect the discriminant analysis was being asked to maximally separate the foregone conclusion that had been submitted to it. If indeed Neumann has done so, this actually would have been experimentally acceptable as long as he had tested all other *a priori* group hypotheses (e.g. Baur's hypothesis, the traditional groupings, etc.) as well. Alternatively, he could have used one of the unsupervised EMVA

³³⁸ Cf. the separate critiques of Mealand and Forbes; Forbes, "Statistical Research on the Bible," 193; Mealand, "The Extent of the Pauline Corpus," 64.

³³⁹ Cf. Borland et al., who comments that discriminant analysis proceeds "on the assumption that the groupings given are correct." Boreland and Galloway, "Authorship, Discrimination, and Clustering," 130. Cf. also Ledger et al., who with even greater clarity state, "This relies on prior information, namely that we know who the authors were....and....which samples belong to which author" (Ledger and Merriam, "Shakespeare, Fletcher, and the Two Noble Kinsmen," 240). It is for this reason that Mealand prefaced his use of MDA with cluster analysis (Mealand, "The Extent of the Pauline Corpus," 61).

methods mentioned above to provide an independent line of evidence to confirm his findings. As far as I can tell, this was not done.

Some may view our critique of Neumann as overly picayune. Indeed, once all is said and done, and more linguistically informed measures are also tested, Neumann's data may indeed be found to have sketched some of the outlines of the complex covariance structure in texts (authorial, genre, cultural, etc.). Our point, however, is that Neumann has left himself open to the critiques that problematize the easy acceptance of his findings.

On balance, however, two overridingly positive things must be said about Neumann's still significant contribution. First, he moved our discipline forward by exercising a more informed experimental design than had been executed prior to his work. Second, at the most coarse level it clearly appears that in pursuing authorship, Neumann discovered, as Burrows and Biber³⁴⁰ did before him, that genre, colloquially said, seems to get in the way. If Radday was the first to definitively discover through the use of multivariate analysis that genre appears to constitute the first major component of variation in the Hebrew Bible, then Neumann appears to have done so for the texts of the NT.

3.10.3.3 H. H. Greenwood and Various Authorship Mixes in the GNT

Late in his career H. H. Greenwood, a quantum chemist at the University of Keele, put his prodigious quantitative skills to work in three articles that explored Pauline authorship and authorship in Luke and Acts.³⁴¹ Despite the originality of his work (through the use of nonlinear mapping or NLM, see below) this triad of articles evidences a number of distinct flaws. (1) Only

³⁴⁰ Biber labels the first factor (which is analogous to the first dimension in discriminant analysis) "high informational density and exact informational content versus affective, interactional, and generalized content." Biber, *Variation Across Speech and Writing*, 107.

³⁴¹ Greenwood, "St Paul Revisited: A Computational Result," *passim*; Greenwood, "Saint Paul Revisited: Word Clusters," *passim* and Greenwood, "Common Word Frequencies and Authorship," *passim*.

high-frequency words were analyzed and in the case of “St [sic] Paul Revisited” only ten were used to develop the analysis. While high-frequency tokens are certainly a viable strategy for feature set selection, they should be contrasted with results developed using FSE-based methods of feature set selection. This was not done. (2) The use of high-frequency words, especially very limited numbers of them, severely underweights minor genres. For this reason adjusted residual-based measures of feature set selection are indicated.^[q] This too was not done.³⁴² (3)

Greenwood’s blanket presumption “that clusters and groups [of texts]...are taken to reflect authorship,”³⁴³ is quite difficult to understand given that he was writing eight years after Burrows and Biber had demonstrated genre as the first component in EMVA work in mixed-genre corpora. (3) Greenwood’s separations, especially in his earlier articles, were simply not very convincing.³⁴⁴ (4) No overall goodness measure (in the case of NLM-like projections it is typically a stress measure) is mentioned. Accordingly, it is difficult to evaluate the usefulness of Greenwood’s reduced space projection.^[r] (5) Perhaps most importantly, NLM (also called Sammon mapping)³⁴⁵ resists perspicuous interpretation because the successive dimensions cannot be represented as a linear combination of the original variables. It is likely for this reason that none have followed Greenwood in using NLM visualizations.

Despite the pointedness of this critique, Greenwood’s last article made noteworthy improvements by using more tokens. He uncovered chapter clusters (he used chapters rather than books) that are conformationally similar to the text conformations presented in Chapter Eight. Specifically, a Johannine complex was cleanly separated from the rest, there was a Colossians-

³⁴² Specifically, if a genre is found in low frequency in the GNT (e.g. paraenesis, Testament, or *Mandata Principis*) the lexical stock that defines that genre will necessarily be in low frequency in the GNT. For methods that use high-frequency tokens (such as Greenwood’s) the overall effect will be to de-weight these minor genres.

³⁴³ Greenwood, “Common Word Frequencies and Authorship,” 184.

³⁴⁴ His figures 2 and 3 show such interpenetration of the chapters as to require interpretive circles and even these are not too convincing (Greenwood, “St Paul Revisited: A Computational Result,” 45).

³⁴⁵ See Sammon, “A Nonlinear Mapping for Data Structure Analysis,” *passim*.

Ephesians pairing, and (perhaps) a *Mandata Principis* (1 Timothy and Titus) pairing.³⁴⁶

Accordingly, if corrected by the above recommendations, NLM or, better yet, the more updated forms of multidimensional scaling (MDS) may indeed demonstrate value in supplementing eigen-systems based EMVA approaches.

3.10.3.4 Gerard Ledger and the Pauline Epistles

At first glance, several aspects of Ledger's effort in his 1995 LLC article seem fairly pedestrian such as his (a) *ad hoc* and somewhat idiosyncratic selection of markers that reflected neither the systems or structures of the Koine³⁴⁷ and (b) his somewhat standard approach of cluster analysis followed by various forms of EMVA.³⁴⁸ One signal experimental design choice, however, sets Ledger's work apart from the other studies explored here. Specifically, Ledger explored the epistles *alone* abstracted from potentially confounding effects from other genres such as Gospel, history/historiography (Acts) and apocalyptic (Revelation). Next, he added those non-epistle texts back in and re-ran the analysis. Quite intuitively, the mixed-genre work demonstrated significantly better separations between text clusters whereas executing EMVA on the epistle texts alone showed little to no observable clusters.³⁴⁹ His contribution pointed toward, though did not explore, the reasonable next step of maximizing the intra-epistolary separation (maximizing the distance between clusters within the epistles and minimizing spread within

³⁴⁶ Greenwood, "Common Word Frequencies and Authorship," 185.

³⁴⁷ Greenwood created a kind of linguistic cocktail with three categories of measures, nineteen ALETS, nine BLETs and the now seldom used Type-Token ratio. The former two measures are creative but idiosyncratic: the percentage of words in each text containing a specified letter of the alphabet, and a specified letter as the ultimate letter in a word (Ledger, "An Exploration of Differences in the Pauline Epistles," 86).

³⁴⁸ Specifically, PCA and discriminant analysis. See Ledger, "An Exploration of Differences in the Pauline Epistles," 86–92.

³⁴⁹ Ledger, "An Exploration of Differences in the Pauline Epistles," 88. To Ledger's credit he recognized in discriminant analysis, and in all supervised approaches, what we referred to as "stacking the deck." He termed the technique "not entirely neutral" (Ledger, "An Exploration of Differences in the Pauline Epistles," 89) but then, unfortunately, extensively used it.

those text clusters) by methods of FSE. This is an indicated “next-step” strategy moving forward.^[s]

3.10.4 Design Implications

The problematic practices of this period can be distilled down to three:

1. Increased number of language measures without any underlying linguistic theory. This might be termed the bag of measures approach.³⁵⁰
2. Increased multivariate techniques but again without any underlying linguistic theory to either drive or interpret the results. This, in turn, might be called the bag of analytics approach.³⁵¹
3. Increased use of multivariate techniques without a theoretical model to interpret that variation.³⁵²

These problems suggest three implications.

Implication 1: Eliminating Type I Measurement Error. Given that we “can’t measure everything” (an enduring chestnut that is now far less true than it once was),³⁵³ we must develop a principled method to down-select the measures that best represent the whole distribution. Simply said, the subset selected must minimize the number of total measures while maximizing the fit between the representative subset chosen and the entire set. The sequence should be as

³⁵⁰ Cooper and Pearsall, “The Gawain Poems: A Statistical Approach,” 374–84.

³⁵¹ Although Peng and Hengartner used both PCA and canonical discriminant analysis to explore the multivariate differences between sixteenth- through twentieth-century authors, all their analytics was predicated on only one category of measures: function words (Peng and Hengartner, “Quantitative Analysis of Literary Styles,” 177). That this critique is not fatal, see the excellent discrimination between Baum and Thompson’s books of Oz in Binongo, “Who Wrote the 15th Book of Oz,” 13–16.

³⁵² McColly serves as an example. Even limited by a relatively small set of function words analyzed only via chi-square, McColly can state that “the highly significant differences in function word frequencies between parts A and B imply that modality is at least as important as authorship in determining the choice of such words.” McColly, “Style and Structure,” 175.

³⁵³ Forbes comments “When the target population is accessible...*then the sampled population can be the target population*” (emphasis mine). Forbes, “A Critique of Statistical Approaches to the Isaiah Authorship Problem,” 533.

follows: (1) understand the layers of language, (2) disentangle them, and (3) identify which measures are disproportionately representative of which layer of variation.

Implication 2: Eliminating Type II Causal Error. This is coincident with Biber's epistemological question related to how one would go about validating hypothesized factor interpretations.³⁵⁴ Biber, importantly, proposes one answer to his own quandary: *confirmatory factor analysis*. Indeed, this is one excellent answer—but it also evidences a certain unsatisfying epistemic circularity—this is empiricism answering empiricism. It seems best to posit that another parallel approach should be executed as well: identification of factors inspired from cross-disciplinary locales such as literary theory, linguistics, genre criticism, rhetorical criticism, and the like. These are disciplines that have busied themselves with explanations for the “communicative functions”³⁵⁵ of language. For this reason top-down experimental designs as well as bottom-up confirmatory designs are indicated.³⁵⁶ Accordingly, both designs are incorporated in the overall design articulated in Chapter Four.

Implication 3: Using Multiple Mathematical and Validation Methods. This translates to using multiple unsupervised EMVA methods as well as multiple supervised methods. This praxis will serve to compare and contrast the derived factors as well as assay their coherence.

3.11 Period Six: The Mess (The 1990s–Mid-2000s)

Even though supervised techniques³⁵⁷ had begun to emerge by the early 1990s replete with FSE techniques in tow, all was still not well in the computational stylistic kingdom. By the middle of the 1990s the seemingly isolated problems in computational stylistics in general and

³⁵⁴ Biber, *Variation Across Speech and Writing*, 91–93.

³⁵⁵ Biber lists eight such functions: (1) participant roles and characteristics, (2) relations among the participants, (3) setting, (4) topic, (5) purpose, (6) social evaluation, (7) relations of participants to the text, and (8) channel. Biber, *Variation Across Speech and Writing*, 29.

³⁵⁶ Either approach places the interpretation of factors on more solid interpretive footing than merely naming a complex according to one's fancy. For an example of this propensity see Guerin-Pace, “Textual Statistics,” 81.

³⁵⁷ See section 3.12.

authorship attribution studies in particular came to a head. It wasn't merely that the Smith/Merriam and Elliott/Foster debates had taken their toll. Nor was it that computational stylistics lacked (extending the Wild West metaphor) new mathematical horses to ride. Nor did the field lack introspective examinations of its collective malaise.³⁵⁸ The problem was that the discipline was really not yet a discipline. Reviewers such as Holmes, Forbes, Parunak, Neumann, Rudman, O'Donnell, Delcourt, Craig, Tuldava, and Juola,³⁵⁹ in particular, began to suggest or state outright that the discipline *itself* was deeply flawed. Rudman, for instance, famously concluded that there was "more wrong with authorship attribution studies than what was right."³⁶⁰ Juola simply pronounced the field "a mess."³⁶¹ His assessment was twofold: the state of authorship attribution was dire, and authorship attribution, in particular, would fail to pass either the *Frye* or *Daubert* tests of scientific validity in courts of law.³⁶² One of two outcomes now seemed certain. Either some elixir would need to be administered to the now centenarian discipline—and quickly—or its own failure to thrive would produce a natural end to one of the longest failed experiments in literary history. We organize our assessment below into four

³⁵⁸ By our count forty-one of the fifty reviews of the stylometric literature were, in fact, published prior to the end of this period.

³⁵⁹ Holmes, "The Analysis of Literary Style," 328–39; Neumann, *The Authenticity of the Pauline Epistles*, 23–114; Rudman, "The State of Authorship Attribution Studies," 352–59; Forbes, "Statistical Research on the Bible," 185–206; Parunak, "Computers and Biblical Studies," 1112–24; O'Donnell, "Linguistic Fingerprints," 207–26; Delcourt, "Stylometry," 979–86; Craig, "Stylistic Analysis and Authorship Studies," 277–84; Tuldava, "The Development of Statistical Stylistics," 145–48; Juola et al., "A Prototype for Authorship Attribution Studies," 170–74.

³⁶⁰ Rudman, "The State of Authorship Attribution Studies," 351. Cf. also Juola, "Authorship Attribution for Electronic Documents," 121. What was interesting about this assessment, of course, was its source. It didn't come from the always present neo-Luddite factions, those scholars who react violently to the encroachment of new methods into their otherwise pacific academic realms of endeavor. It didn't come from Schoenbaum who long ago critiqued the whole exercise as "the bedlamite antics of the wildmen" (Schoenbaum, *Internal Evidence and Elizabethan Dramatic Authorship*, 107) or from Vickers who viewed it as a somewhat wayward exercise depending on an "atomistic notion of style." Rather, it came from the discipline's own practitioners themselves.

³⁶¹ Juola, in particular, had already cemented his role as a main figure in cleaning up the mess by sponsoring the ALLC/ACH "Ad-hoc Authorship Competition" in 2004. Juola, "Authorship Attribution for Electronic Documents," 123–26. Nonetheless, by 2006 Juola could still accurately summarize the state of authorship attribution studies as follows: "Despite a century of research, statistical and computational methods for authorship attribution are neither reliable, well-regarded, widely used, or well understood" Juola et al., "A Prototype for Authorship Attribution Studies," 169.

³⁶² Juola et al., "A Prototype for Authorship Attribution Studies," 170–71.

categories: the mess in findings, the mess in measures, the mess in methods, and the mess in the lack of an overarching framework (or experimental design).

3.11.1 The Mess Defined

3.11.1.1 The Mess in Contradictory Findings

By the time Juola and Rudman wrote their separate reviews, contradictory findings in computational stylistics were nothing new.³⁶³ The problems were so pandemic, in fact, that a frustrated Rudman collected his now famous list of eleven problems in the discipline.³⁶⁴ In our own discipline of Pauline studies we chronicled (per Section 2.3.3) contradictions in which Morton asserted that Paul wrote only four epistles (the *Hauptbriefe*), Grayston and Herdan concluded ten (all epistles but the Pastorals),³⁶⁵ Gerard Ledger assigned six to Paul, Barr concluded six or seven, and Kenny concluded that twelve were “the work of a single, unusually versatile author.”³⁶⁶

3.11.1.2 The Mess in Linguistic Measures

Our meta-analysis of the literature (see Section 2.4.3) demonstrated that by the beginning of the 1990s no fewer than sixty-three categories (types) of linguistic measures had been deployed in authorship studies, covering tens of thousands of individual tokens.³⁶⁷ Yet despite

³⁶³ Memorably, as early as 1971 Tallentire concluded, “The conclusions reached about style, via computational methodologies, far from being ‘definitive’ by virtue of their objectivity, are mutually contradictory.” Tallentire, “Mathematical Modelling in Stylistics: Its Extent and General Limitations,” 117. Mealand, more sanguine than Tallentire, still registered reservations about the discipline characterizing it as “a still developing subject ... [with] ... no stable consensus about the criteria for testing style.” See Mealand, “Computers in New Testament Research,” 106–10.

³⁶⁴ Rudman, “The State of Authorship Attribution Studies,” 352–53.

³⁶⁵ To be fair, Grayston and Herdan didn’t exactly say Paul, but “a first century writer,” who was presumably Paul. See the footnote immediately below.

³⁶⁶ Grayston and Herdan, “Authorship of the Pastorals,” 13–14; Morton, “The Authorship of Greek Prose,” 224–25; Barr, *Scalometry and the Pauline Epistles*, 125; Ledger, “An Exploration of Differences in the Pauline Epistles,” 95; Kenny, *A Stylometric Study of the New Testament*, 100.

³⁶⁷ The categories are as follows: Adjectives, Adverbs, Articles, Bigrams, Bag of Parts of Speech, Bag of Words, Bag of “Various Kinds” of measures, Capitalization, Clause Types, Collocation and Colligations, Conjunctions, Diminutives, Direct Address, Distance between Words or Stopwords, Entropy of Words or Syllables, First Words of Speeches, Function Words, Genre, *Hapax Legomena*, Idioms (Figures of Speech), Infinitives, Letters, Lexical Tokens, Metrical Evidence (Metre-related), Order, Other Language (e.g. Latinisms), Polarity, Particles, Parts of

the prodigious amount of work this entailed, virtually all of this work was performed without an underlying linguistic theory other than traditional grammar.^[a] By the mid-1990s principled work using linguistically-informed measures to disentangle idiolectic from sociolectic layers in texts was still conspicuous for its absence.

3.11.1.3 The Mess in Quantitative Methods

A similar meta-analysis demonstrates that by the same time (the beginning of the 1990s) no fewer than forty-one kinds of mathematical methods had been used in authorship studies.³⁶⁸ As above, no consensus existed on the relative value or sequence of each method toward understanding linguistic phenomena, nor its various layers, nor the extent of its presumed latent factors.

3.11.1.4 The Mess in Experimental Design

In 1992 Clayman aptly described the mess in experimental design with these words:

The intelligent choice and handling of statistical measures depends on two sets of assumptions: assumptions about the data made by each statistical test and assumptions about language, literature, and the act of writing that underlie the philological or historical point at issue. The importance of the latter is illustrated clearly by chronological and authorship studies. These were the earliest stylometric...studies...and remain the most problematic. They proceed in a way exactly opposite to LASLA, beginning with a thesis and admitting only data and statistical analyses that appear to prove it.³⁶⁹

Speech, Phonemes, Pronouns, Proportionate Pairs, Punctuation, Prepositions, Rare Words, Repetition, Roots or Stems, Semantic Groups, Sentence Endings, Sentence Length, Sentence-Length Variability, Special-Character, Stop Words, Syllables, Syntactical Tokens, Taxis, Tenses, Theme, Traditional Grammar, Transitivity and Ergativity, Trigrams, Verbs, Relative Frequency, Vocabulary, Voice, Word Groups, Word Positions, Words High Frequency, Word Length, Word-Links, Words Low Frequency, Words Selected, Words Shared, Words (Key Word in Context).

³⁶⁸ The analyses include, without limitation, ANOVA, Bayesian Multinomial Regression, Bayes Theorem, Characteristic Curves, Cluster Analysis, Coefficient of Variation, Correlation, Correspondence Analysis, Cross-Entropy, Decision Tree Approaches, Descriptive Stats (Non-inferential), Discriminant Analysis, Distance (Euclid, Mahal, Ziv-Merhav etc.), Exact Tests, Factor Analysis, Frequencies or Normalized Frequencies, Negative Binary Distributions, Full Bayes Poisson, Information Theory (Entropy), Karhunen-Loeve-Transform, K-Means, Kruskal-Wallis Test, Likelihood Ratio, Lempel and Ziv approach, Markov Models, Multiple Regression, Negative-Binomial, Non-Statistical Tabular Comparison, Principle Components Analysis, Poisson Distribution, Statistical Inference: Multivariate, Statistical Inference: Univariate (Chi-square/t-test/z), Stat. Inference: Unusual (Monte Carlo), SVM: Support Vector Machine, Visualization (Descriptive), Visualization Multivariate, Word Recurrence Interval, Yule's K, Zipf's Law, and "Other."

³⁶⁹ Clayman, "Trends and Issues in Quantitative Stylistics," 387.

Clayman makes two salient points related to experimental design, one declarative and one by inference. Regarding the first, although he overplays his hand (stylometric studies by and large haven't typically admitted "only data and...analyses that appear to prove it"), these studies *have* clearly picked narrow sets of markers *ad hoc*.^[b] Second, by inference, most studies have not been properly inductivist at all. Stylometricians have not first of all sought to describe language-in-use, and from that copious level of description sought to develop principles of language subset selection up from the data itself.

3.11.2 Understanding the Mess through Kuhnian Lenses

How, then, might we interpret all the dimensions of this mess? Without question, stylometry up to the time of the "mess" was a collision of storm fronts. In one tangled complex were studies that measured all items vs. others that picked markers *ad hoc*, where some research was univariate and other multivariate, where priors were stated or not, where procedures were clear or unarticulated, where validation was present, unclear...or absent. Intriguingly, this complex of "messes" appears to be an almost textbook example of Kuhn's crisis/novelty paradigm described in his seminal 1962 work, *The Structure of Scientific Revolutions*.³⁷⁰ In hindsight, we can see now that by the middle of the first decade of the new millennium, a Kuhnian paradigm shift had occurred—but few caught in the mess had gained enough elevation above it to recognize that the discipline was in fact beginning to slog its way out.

3.11.3 The Design Implications from the Mess

The design implications are rich here. Rudman lists six³⁷¹ with the most intriguing of these being his call for more theoretical grounding in the discipline. Moreover his

³⁷⁰ "Crises are a necessary precondition for the emergence of novel theories." Kuhn, *The Structure of Scientific Revolutions*, 77.

³⁷¹ Rudman, "The State of Authorship Attribution Studies," 361.

recommendation to “study style in its totality”^[c] was both a rebuke for the older *ad hoc* pick-a-marker praxis as well as a recognition that language has more to say if we permit it to do the talking. To this we may add his stress on reducing error.^{372[d]} Last, we would stress that this period writes one final lesson large upon our praxis: the need to embrace both unsupervised and supervised approaches for the two distinct and synergistic outcomes they respectively bring to computational stylistics, that of understanding the fundamental structure of style on the one hand and classification on the other.^[e]

3.11.4 Slogging Out of the Mess (1995-Current)

It would overreach the history of the discipline to claim that in response to the “mess” researchers embraced a coordinated improvement program, say, by following Rudman’s prescriptions or taking Forbes’ critiques to heart.³⁷³ Despite the absence of such a program, several notable themes discursively distilled from the collectively-felt *angst* of the 1990s: deeper measures, deeper organization, the birth of corpus linguistics and the veritable explosion of methods of supervised mathematics applied to classification tasks in computational stylistics.

First, in terms of deeper measures, *studies began to be developed first with dozens and then hundreds of separate measures—and then millions of combinations of the same.*^[e] Lana, as we saw previously, used the entire lexical stock of 30,745 words in his study of Xenophon,³⁷⁴ and Dixon and Mannion tested 7,726,160 combinations of their twenty-six measures taken fifteen at a time.³⁷⁵ These kinds of studies, made tractable by advances in low-cost computing, signaled a new era in measurement adequacy within computational stylistics.

³⁷² Rudman, “The State of Authorship Attribution Studies,” 358.

³⁷³ Forbes, “A Critique of Statistical Approaches to the Isaiah Authorship Problem,” 531–40.

³⁷⁴ Lana, “Xenophon’s Athenaion Politeia,” 22.

³⁷⁵ Dixon and Mannion, “Goldsmith and the Public Ledger,” 310.

Second, objective competitions to discover best practices began to be held.³⁷⁶ When a far more comprehensive history of computational stylistics is written it may indeed record that the ALLC/ACH-hosted competition held in June 2004 was an inflection point—the point at which the mess in traditional authorship attribution was already beginning to be cleared just as it was being pronounced its most extreme. “Results from this competition were heartening,” wrote Juola, and were even “unbelievable,” in the enthusiastic words of one contest participant.³⁷⁷ At the 2004 competition, for the first time a number of research teams successfully navigated a difficult data set of thirteen disparate problems. One team, Koppel and Schler, scored a quite remarkable average success rate of seventy-one percent correct classification.³⁷⁸

Third, researchers began the process of the *principled collection of texts* as the first step in computational stylistics. This, in turn helped foster an already emergent discipline (or is it a subdiscipline?) termed corpus linguistics. Its early advocates included many of the corpus stylistics researchers we have already met in these pages.³⁷⁹

Fourth, and perhaps most importantly, *a period of retooling and constrained successes* emerged during this time. This turn for the better was reflected in Burrows’ comment as early as his 2001 Roberto Busa lecture, “We have not escaped the battles that so excited the media....but our methods are increasingly reliable, our use of them is ever more rigorous, and we have vast new corpora to strengthen our comparisons.”³⁸⁰

³⁷⁶ In the commercial data mining arena similar annual competitions are held, with the main competition being the KDD cup held by the ACM Special Interest Group on Knowledge Discovery and Data Mining.

³⁷⁷ Juola et al., “A Prototype for Authorship Attribution Studies,” 172.

³⁷⁸ Juola et al., “A Prototype for Authorship Attribution Studies,” 172.

³⁷⁹ See especially Biber’s monograph (Biber et al., *Corpus Linguistics*, 243–80). Cf. Forsyth and Holmes, “Feature-Finding for Text Classification,” 163–64; Baayen et al., “An Experiment in Authorship Attribution,” 30.

³⁸⁰ Burrows, “Questions of Authorship,” 7.

3.11.5 Design Implications from Slogging Out of the Mess

The primary findings discussed at the 2004 ALLC/ACH authorship attribution competition can serve to summarize the kinds of designs we should adopt in our own abductive effort here:

1. *Use measures across the scale of rank.* The conference called this “syntax priority,” stating that “methods based on simple lexical statistics tended to perform substantially worse than methods based on N-grams or similar measures of syntax in conjunction with lexical statistics.”^{381[f]}
2. *Use supervised multivariate methods* as the vanguard of a new wave: “The best methods were multivariate with the leading method being a support vector machine that down selects language measures to those that estimate the population of linguistic measures.”³⁸²
3. *Implement software automation:* Juola lays out the glimpse into the future, the possibility of automated authorship attribution with “off-the-shelf” software.^{383[g]}

3.12 Period Seven: The New Stylistics 2.0: Machine Learning (Mid-1990s–Current)

If the first salvo of the New Stylistics was fired by *unsupervised extractive multivariate analysis* (EMVA) the second salvo was fired in the early 1990s via *supervised machine learning methods*. In the primer on computational stylistics in chapter two we clarified that supervised approaches are, in reality, three separate technologies: (1) automated FSE,³⁸⁴ (2) advanced algorithms such as machine-learning (ML) techniques, all operating within (3) a closed loop validation system. But this hardly captures the full ethos of the matter, for in the span of the last twenty years supervised methods not only emerged as the dominant arm of the New Stylistics but birthed the 10-billion-dollar business intelligence industry.³⁸⁵ The story, however, doesn’t

³⁸¹ Juola et al., “A Prototype for Authorship Attribution Studies,” 172.

³⁸² Juola et al., “A Prototype for Authorship Attribution Studies,” 172.

³⁸³ Juola et al., “A Prototype for Authorship Attribution Studies,” 173–76.

³⁸⁴ For a deeper overview of FSE techniques, see Mukherjee and Liu, “Improving Gender Classification of Blog Authors,” 211. For the differences between the filter approach and wrapper approaches see Blum and Langley, “Selection of Relevant Features,” 245–71 and Kohavi and John, “Wrappers for Feature Subset Selection,” 273–324, respectively.

³⁸⁵ Not in academia (!) but in the commercial knowledge discovery, data mining and business intelligence markets worldwide.

end there. The last ten years, in particular, have witnessed what can only be described as an explosion within computational stylistics due to advances in supervised classification. Prior to the mid-2000s there were only one-tenth the useful phyla of stylometric approaches as there are now. This period witnessed—for the first time—the long-awaited arrival of both the necessary mathematics and a closed-loop cycle methodology sufficient for solving broad classes of classification problems in computational statistics. To address this era in detail would go far beyond our current scope. What we can detail briefly is that two separate factors converged to produce this “explosion,” and that our experimental design can benefit from both:

1. *Machine Learning Rapidly Matured and Diversified.* A machine learning method is a mathematical method that learns from the data that it is given (say hundreds or thousands of linguistic features in common between dozens of texts) and discovers which of these linguistic measures, in what combinations and in what layers (features feeding into features) best classify known texts. The following high-level factors certainly contributed to the remarkably quick adoption of supervised ML methods:
 - a. Early adopters of ML methods quickly reaped higher levels of success than with prior approaches. These early adopters included Tweedie, Singh, & Holmes (1996); Lowe & Matthews (1995); Khmelev & Tweedie (2001); Diederich, Kindermann, Leopold, & Paass (2000); and Argamon, et al. (2003).³⁸⁶

³⁸⁶ Tweedie et al., “Neural Network Applications in Stylometry,” passim; Lowe and Matthews, “Shakespeare vs. Fletcher,” passim; Diederich et al., “Authorship Attribution with Support Vector Machines,” passim; Argamon et al., “Gender, Genre, and Writing Style,” passim.

- b. Many supervised learning techniques came online quickly, beginning with neural networks,³⁸⁷ followed by genetic algorithms,³⁸⁸ support vector machines (SVM), and then ensemble classifiers (weighted voting).³⁸⁹
 - c. SVM's, in particular, quickly carved out a niche in handling high-dimensional feature spaces. Pioneers in the application of SVM included Diederich et al. (2003); de Vel et al. (2001); Zheng, and Qin, Huang, & Chen (2003).³⁹⁰ Time would demonstrate that quite often SVM's seem to outperform most other MLT's.³⁹¹
 - d. Initial efficient results were generated with these methods operating upon N-grams, which are letter frequency or letter pairs,³⁹² sometimes with even as few as five letters.³⁹³
2. *The Rapid Maturation of a Machine Learning Methodology.* This is the three-fold feedback loop we mentioned above, an automated ballet in which machine-learning methods operate in concert with FSE and multiple methods of validation.

For a more complete discussion of these methods, including ANN's (Artificial Neural networks),³⁹⁴ PNN's (Probabilistic Neural Networks),³⁹⁵ SVM's, and Bayesian regression, see

³⁸⁷ For introductions to neural networks see Matthews and Merriam, "Neural Computation in Stylometry I," 203–05; Lowe and Matthews, "Shakespeare vs. Fletcher," 450–54. For a more mathematical (but still generally accessible treatment) see Tweedie et al., "Neural Network Applications in Stylometry," 2–3; Tweedie, "A Statistical Investigation into the Provenance of De Doctrina Christiana," 21–32.

³⁸⁸ See Holmes' helpful summary in Holmes and Forsyth, "The Federalist Revisited," 121–22.

³⁸⁹ See especially Dietterich, "Ensemble Methods in Machine Learning," passim; Stamatatos, "Authorship Attribution Based on Feature Set," passim; Preisach and Schmidt-Thieme, "Ensembles of Relational Classifiers," passim.

³⁹⁰ See Diederich et al., "Authorship Attribution with Support Vector Machines," passim; de Vel et al., "Mining E-Mail Content," passim; Zheng et al., "Authorship Analysis in Cybercrime," passim.

³⁹¹ See especially de Vel et al., "Mining E-Mail Content," 58; Dewdney et al., "The Form is the Substance: Classification of Genres in Text," 7.

³⁹² Kjell, "Authorship Determination Using Letter Pair Frequency," 120; Kjell, "Authorship Attribution of Text Samples," 1660–61; Merriam, "Letter Frequency as a Discriminator of Authors," 468; Merriam, "Heterogeneous Authorship in Early Shakespeare," 18–19; Hoorn et al., "Neural Network Identification of Poets," 314–15.

³⁹³ Merriam, "Edward III," 182.

³⁹⁴ Waugh et al., "Computational Stylistics Using Artificial Neural Networks," 187–90.

³⁹⁵ Martindale and McKenzie, "On the Utility of Content Analysis," 262–63.

Koppel et al. (2010).³⁹⁶ Juola, Stamatatos, and Argamon³⁹⁷ provide helpful summaries of this era and its new methods.

3.12.1 Design Implications

The new era of ultra-high performance supervised ML methods produced a number of findings that resulted in best practices. In particular these best practices (which were quickly disseminated) yielded the current modern levels of excellent performance, often above ninety percent correct classification in the hold-back sample in *ideal authorship attribution* (IAA) problems.³⁹⁸ These best practices included:

- Use of an adequate training set³⁹⁹
- Use of multiple ML methods
- Use of multiple validators
- Use of large “vectors” of N-grams or other lexically-based linguistic features

One final design implication should be discussed. Researchers discovered in this era that supervised methods can quite easily produce models and results that are *overfit*. These kinds of solutions are nongeneralizable to other author sets,⁴⁰⁰ and even less generalizable to differences of sociolect (genre, dating, etc.). In fact, all such supervised solutions, overfit or not, are non-generalizable across causal boundaries precisely *because supervised methods can only be trained on one outcome at a time*. Therefore, despite their efficiency at whatever single objective they

³⁹⁶ Koppel et al., “Computational Methods in Authorship Attribution,” 11.

³⁹⁷ Juola, “Author Attribution,” passim; Koppel et al., “Computational Methods in Authorship Attribution,” passim; Stamatatos, “A Survey of Modern Authorship Attribution Methods,” passim; Argamon and Koppel, “Finding Meaning in Stylistic Variation,” passim. For a less technical introduction to modern authorship attribution and its leading researchers see Calle-Martin and Miranda-Garcia, “Stylometry and Authorship Attribution,” passim.

³⁹⁸ It should also be recognized that modern supervised methods operating within one genre and operating upon an adequate number of linguistic markers achieve similar levels of discrimination (Burrows and Craig, “Authors and Characters,” 293–306).

³⁹⁹ Hoover and Hess, “Authorship Attribution: the Mysterious Maria Ward,” 467.

⁴⁰⁰ That is, when one attempts to use those same experimentally derived markers on other sets of authors, they frequently fail to yield an efficient discrimination.

are trained upon, this very training to a single objective makes them wholly inappropriate for arriving at an understanding of multiple causation.^[h]

3.13 Milestone Five: Inviting Linguistics to the Dance (ca. 2005–Current)

In the various subsections of this chapter I have repeatedly underscored the need for a more linguistically informed approach to texts, paradigmatically, syntagmatically, stratally, and pragmatically.^[a] Nonetheless, the linguistic choices made by many researchers today are still idiosyncratic and lack any discernable relation to mainstream linguistic theory.⁴⁰¹ In response to this, Argamon, following Craig,⁴⁰² has called for a more *linguistic approach* to computational stylistics: “In the long term, however, a clear foundation in a linguistic theory of meaning will be needed to gain true insight into the nature of the stylistic dimension(s) under study.”⁴⁰³

Accordingly, a small cadre of researchers have begun to annotate corpora using categories explicitly drawn from linguistic theory. This very new development seems to have centered on SFL-based studies.^[b] These efforts include the work of Stamatatos, Fakotakis and Kokkinakis (2001) who have explored the scale of rank in Modern Greek,⁴⁰⁴ Whitelaw’s (2006)⁴⁰⁵ study of interpersonal distance, Argamon’s (2007)⁴⁰⁶ textual classification study, and Nini’s authorship attribution research (2013).⁴⁰⁷ Mention too, should be made of the entire subdiscipline of Corpus Linguistics and in particular the quite helpful applied work of Matthew O’Donnell in terms of the Greek New Testament.⁴⁰⁸ Beyond this modest amount of research, I am aware of no other

⁴⁰¹ Thus Sapkota et al.’s work which used “stylistic, lexical, perplexity values from character level N-gram language models, and syntactic features.” Sapkota et al., “The Use of Orthogonal Similarity Relations,” 465.

⁴⁰² “Stylistics,” Craig stated, “is perhaps best seen as the offspring of Saussurean linguistics.” Craig, “Stylistic Analysis and Authorship Studies,” 279.

⁴⁰³ Argamon and Koppel, “Finding Meaning in Stylistic Variation,” 81.

⁴⁰⁴ Stamatatos et al., “Computer-Based Authorship Attribution,” 199–201.

⁴⁰⁵ Whitelaw et al., “Identifying Interpersonal Distance,” 200–12.

⁴⁰⁶ Argamon et al., “Stylistic Text Classification,” 803–17.

⁴⁰⁷ Nini and Grant, “Bridging the Gap between Stylistic and Cognitive Approaches to Authorship,” 177–98.

⁴⁰⁸ O’Donnell’s work (O’Donnell, *Corpus Linguistics*) is particularly important in two regards, First it is grounded in a mature understanding of SFL theory, and he has used it to rethink standard approaches to New Testament problems including textual and source criticism, lexicography, and discourse analysis. Beyond O’Donnell, however,

large-scale recent linguistically-informed research that has embraced other language schemata outside of SFL.⁴⁰⁹

3.13.1 Design Implications

In our view this newly emergent “linguistic turn” is both timely and necessary. We submit that two further design implications should be considered:

1. The SFL measures developed for our purposes (disentangling genre from authorship) should include multiple strata, syntagmatic structures, paradigmatic systems, and metafunctions of the GNT.
2. Given the absence of other potentially adequate functional systems, abduction requires that we test other systems for their utility as well. We suggest as candidates Dik’s functional grammar, Lamb’s stratificational grammar and Fawcett’s system of syntax.^[c]

much recent work in Corpus Linguistics, seems to lack a linguistic apparatus, or quantitative currency. Biber, in my view is a notable exception. See especially Biber et al., *Corpus Linguistics*; Biber, *University Language*; Biber et al., *Discourse on the Move*.

⁴⁰⁹ Linmans mentioned Dik but only in passing (Linmans, “Correspondence Analysis of the Synoptic Gospels,” 13).

4 Developing the Design for Disentangling Authorship from Genre

This is the fulcrum chapter of this work. The goal here is to leverage the findings and implications from the prior chapter in order to develop a *comprehensive and adequate* experimental design for our central task: disentangling authorship from genre in the GNT. To achieve something that approaches *comprehensiveness and adequacy*, however, two final concepts are necessary:

First, the review just completed demonstrated that, especially from 1987 onward, dozens of EMVA studies executed upon mixed-genre corpora *outside the NT* achieved satisfying separations between idiolectic and sociolectic layers in their respective corpora. But in our own central task of disentangling genre from authorship in the mixed genre corpus of the GNT—and here is our point—our experimental design work must *not* be restricted to merely imitating what has been done already. The brand of experimental design we are aiming for here is *synthetic*; it must leverage insights from “what has worked” in the past, to be sure. Yet, it must also avoid the problems that have caused prior designs to yield either suboptimal or even misleading results. It must *synthesize what has been done with what should be done*, all informed by the current state of the research for that particular objective.

Second, we must avoid the foundationalist fiction that an experimental design is some sort of objectivist exercise that will unerringly issue forth the “answer” to a problem. In reality, an experimental design is simply a skeleton upon which we hang a sequence of experiments. If we have “done our homework” that experimental design will prove itself to be comprehensive and adequate only in retrospect—by allowing the researcher to arrive *at a state of high probability* with respect to an hypothesis *in the least number of abductive cycles*.

I propose that six steps are necessary toward developing a synthetic design capable of arriving at a state of high probability with respect to disentangling authorship from genre:

1. Define the research scope (by borrowing “what has worked” from research outside of NT studies, per chapter three).
2. Assemble implications from the design review (from chapter three).
3. Assemble implications from current large-scale experiments in computational stylistics research.
4. Develop the initial three dimensions of the synthetic experimental design by bringing together the implications from steps 1-3.
5. Explore some innovative enabling technologies to automate the generation and execution of the synthetic experimental design.
6. Select from that larger design the parts we will execute in chapters five through nine—the inaugural loop of the abductive research cycle.

4.1 Define the Research Scope

The first task, defining the research scope, is easy enough. This is simply the central objective of this work (disentangling genre from authorship in the GNT) as informed by any findings achieved outside of the GNT that may be relevant to achieving that same goal when applied to the GNT. The findings achieved outside the GNT that are most pertinent to our goal of disentangling authorship from genre are as follows:

1. The early and oft-repeated chestnut that *hapax legomena* function as some sort of “fingerprint” or discriminatory marker *par excellence* for authorship has turned out to be mistaken.¹ Subsequent research, in fact, demonstrates that *hapax legomena* in mixed genre texts tend to be more associated with genre than authorship.² To complete the historic reversal, in the 1980s high-frequency content words (the so-called Burrows

¹ See especially Campbell, *Sophistes and Politicus*, xxxi-xxxiii; Cadbury, “Some Foibles of New Testament Scholarship,” 215; Mosteller, “The Writing Styles of the Authors of ‘the Federalist’ Papers,” 140. The central insight that clarifies much of this work is the inescapable fact that special vocabulary that is decidedly not low frequency within a homotypic corpus *will become* lower frequency vocabulary within a heterotypic corpus, especially when a corpus of interest (e.g. the single-audience paraenesis of the Pastorals) is in low prevalence within that heterotypic corpus.

² See especially Greenspahn, “The Number and Distribution of Hapax Legomena,” 18–19. In a homotypic corpus *hapax legomena*, or other low-frequency words may indeed have utility.

method) rather than low-frequency words have proven to be the gold standard for authorship discrimination,³ and since the mid-1990s vectors of the same high-frequency words have proven to be the most successful discriminators in FSE studies.⁴

Implication: Build *hapax* studies into the research by studying authorship and genre by frequency fractions that not only include linguistic categories that occur once, twice, etc., but categories by centile, decile, quartile, and so on. This strategy is called binning.

2. The storied procedure of selecting linguistic features *ad hoc* has yielded only an idiosyncratic disaster.⁵ Its praxis has been overthrown by feature-set extraction (FSE) methods.⁶ *Implication:* Use FSE in both unsupervised and supervised multivariate work toward achieving our core objective.

³ Burrows explanation of the superiority of high-frequency words lies in the fact that they “provide the main strands of the whole fabric.” Burrows, “A Second Opinion,” 377. Dozens of studies have adopted the “Burrows Method” including Forsyth et al., “Investigating the Authenticity of the Consolatio,” 379–82; Hoover, “Multivariate Analysis and the Study of Style Variation,” 341–42.

⁴ Examples include Argamon et al., “Stylistic Text Classification,” 808; Smith and Rickards, “The Authorship of the American Declaration of Independence,” 21; Luyckx, *Scalability Issues*, 28; Kestemont et al., “Robust Rhymes,” 60–62.

⁵ In Argamon’s words, “stylometric models...have typically been based on hand-selected sets of...features.” Argamon et al., “Style Mining of Electronic Messages” 476. These, Forsyth and Holmes lamented, were left “up to the intuition of the analyst.” Forsyth and Holmes, “Feature-Finding for Text Classification,” 163. *Ad hoc* selection has been heavily critiqued per Delcourt, “Statistical Analysis of Co-occurrence,” 21–27; Craig, “Stylistic Analysis and Authorship Studies,” 286–87.

⁶ Automated Feature Set Extraction began with Thisted and Efron, “Did Shakespeare Write a Newly-Discovered Poem?,” 445–55; Valenza, “Are the Thisted-Efron Authorship Tests Valid?,” 28–45 and the practice has completely supplanted *ad hoc* marker selection. In the words of Mosteller and Wallace the hand-picked markers were “annihilated” by machine-picked markers (Mosteller and Wallace, “Inference in an Authorship Problem,” 285).

3. Earlier stylometric studies assigned, *de facto*, significant stylistic variation to authorship alone.⁷ Now, close to 100 studies in traditional textual stylistics⁸ and over 200 overall⁹ demonstrate that not only do other sociolectic/sociocultural sources of stylistic variation exist (i.e. genre, gender, dating, audience, and the like)¹⁰ but subcategories within them can be empirically distinguished as well.¹¹ *Implication:* Use EMVA in the GNT to disentangle genre from authorship.
4. Genre-related findings have also been demonstrated as follows:
 - a. Univariate evidence began to accumulate that, in mixed-genre corpora, genre accounted for a higher proportion of the total univariate stylistic variation than authorship.¹²

⁷ Examples of this are legion. Notable works in biblical studies that make this equivalence despite operating within a mixed-genre corpus, or within single texts with multi-genre include, without limitation: Harrison, *The Problem of the Pastoral Epistles*, 84–86; Morton, “The Authorship of Greek Prose,” 224; Morton, *Literary Detection*, 165–83; Bee, “Statistical Methods in the Study of the Masoretic Text,” 622; Bee, “A Statistical Study of the Sinai Pericope,” 421; Neumann, *The Authenticity of the Pauline Epistles*, 206–22. Also included here is the earlier Radday. See especially Radday, “Isaiah and the Computer,” 73; Radday et al., “The Book of Judges Examined,” 494–99. Craig terms such an assumed assignment the first signal pitfall of computational stylistics (Craig, “Stylistic Analysis and Authorship Studies,” 286).

⁸ By textual stylistics I mean stylistics operating upon written or spoken *language-in-use* texts. These include the various corpora studied by corpus linguists (e.g. the British National Corpus, the Cambridge International Corpus, etc.) or texts collected for historical or other reasons. The GNT is a corpus of this type.

⁹ This is arrived at by adding to studies executed upon standard written and spoken texts those from *extra-textual stylistics*. This includes idiolectic and sociolectic studies on nontraditional forms of language-in-use such as web content, computer source code, social media sources (e.g. Twitter), etc.

¹⁰ Per Wenham, “[What]...emerges from the study [is that] different genres of literature...have distinct styles.” Wenham, “Genesis: An Authorship Study,” 6.

¹¹ Illustratively but not exhaustively, poetry is separated from prose or narrative from dialog in Herdan, *The Advanced Theory of Language as Choice and Chance*, 206–13; Merriam, “Invalidation Reappraised,” 419; Burrows, “The Statistical Analysis of Narrative Style,” 64; Biber, *Variation Across Speech and Writing*, 101–69. Similarly, male is discriminated from female in Argamon et al., “Gender, Genre, and Writing Style,” 326–42; Rustagi et al., “Learning Age and Gender,” 207–11; Cheng et al., “Author Gender Identification from Text,” 80–86. Lastly, authorship is differentiated by date in Temple, “A Multivariate Synthesis,” 69–74; Can and Patton, “Change of Writing Style with Time,” 66–77; HaCohen-Kerner et al., “Stylistic Feature Sets as Classifiers,” 852–59.

¹² Ellegård appears to have been the first to articulate this concern (Ellegård, *A Statistical Method for Determining Authorship*, 39). For studies that demonstrated the confounding effect of genre see Binongo and Smith, “Statistical Approaches to Philippine Literature,” 522; Radday and Shore, “The Definite Article: A Specifying Discriminant,” 26. For studies that concluded that genre stylistic variation was equal to or greater than authorship stylistic variation, see Merriam, “The Authorship of Sir Thomas More,” 104–05; Smith, “Recent Experience and New Developments,” 75; Merriam, “An Investigation of Morton’s Method: A Reply,” 57; Holmes, “Authorship Attribution,” 88; O’Keefe, “Critical Remarks on Houk’s Statistical Analysis,” 424.

- b. Extractive multivariate (EMVA) evidence confirmed these findings. That is, in mixed genre corpora EMVA studies either more frequently extracted genre rather than authorship as their first component *or* the bimodal conformation of the first two factors was interpreted to be more similar to genre.¹³ Initial biblical evidence continues this trend. Thus Neumann writes that his reduced variables “really separate types of writing style more than authors,”¹⁴ and Mealand concludes that more work on genre and authorship is indicated.¹⁵ *Implication: Use EMVA in the GNT to disentangle genre from authorship.*
- c. These dual findings indicate, interpretively, that speakers/writers/redactors vary their style and lexical stock to fit shifting contexts of situation and culture. *When, however, the corpora is more homotypic, authorship variation reappears as the first component of variation.*
- d. In terms of the GNT, however, given that neither the strata, structure, systems nor metafunctions of the GNT have been thoroughly studied, these genre conclusions

¹³ Typically in unsupervised exploratory decompositional data analysis *with mixed genres*, the first component of variation tends to be genre rather than authorship. See Burrows, “The Statistical Analysis of Narrative Style,” 64; Forsyth et al., “Investigating the Authenticity of the Consolatio,” 383; Juola and Baayen, “Authorship Identification by Cross-Entropy,” 63. Cf. Burrows, “The Interpretative Nexus between Analysis and Information,” 92–102. Cf. also Radday et al., “Genesis, Wellhausen and the Computer,” 477–81; Radday and Shore, *Genesis: An Authorship Study*, 100–22; Baayen et al., “An Experiment in Authorship Attribution,” section 3. See also Brainerd who separated Shakespeare’s genre using pronouns (Brainerd, “Pronouns and Genre in Shakespeare’s Drama,” 14–15. From the perspective of SFL this is explicable from the variant deictic profile (a reflection of the interpersonal metafunction) expected between genres.

¹⁴ Neumann, *The Authenticity of the Pauline Epistles*, 184.

¹⁵ After observing the surprising relative proximity of 1 and 2 Peter and the closer proximity of 2 Thessalonians than 1 Thessalonians to Paul (both via Mahalanobis distance), Mealand opined, “In future work I would like to test more specifically...author genre differences.” Mealand, “The Extent of the Pauline Corpus,” 86. While Mealand did indeed return to explore intra-Markan genre, and the genre of Acts and Q within their respective genre sets (see Mealand, “Measuring Genre Differences in Mark,” 227–45; Mealand, “Style, Genre, and Authorship in Acts,” 479–505; Mealand, “Is there Stylometric Evidence for Q?,” 483–507) his original quarry, the NT epistles, lies unexamined.

must be considered tentative at best. *Implication*: Systematic work in all these areas is indicated.

5. Supervised machine-learning methods have become the largest breakout success story in computational stylistics, not infrequently developing classification accuracies above 90 percent¹⁶ in “plain vanilla”¹⁷ attribution tasks. The downside? Because these methods incorporate prior information (the multiple discrete categories of their training sets) they necessarily eliminate the linguistic/stylistic covariance necessary to disentangle the various components of variation in text corpora. Colloquially stated, they throw out the stylistics baby with their supervisory bathwater. *Implication*: Balance supervised with unsupervised multivariate analysis *and* sequence the work so that supervised MVA operates upon the separate fractions obtained from unsupervised MVA.
6. By careful design, both sequential extractive multivariate analysis and supervised approaches (by varying the training outcome) have yielded from texts certain cultural and contextual components other than genre. These have included date,¹⁸ sociolectic

¹⁶ See, for instance, the studies by Abbasi and Chen, “Applying Authorship Analysis to Extremist-Group Web Forum Messages,” 72; Stańczyk, “On Performance of DRSA-ANN Classifier,” 178 and Marsden et al., “Language Individuation and Marker Words,” 7.

¹⁷ An expression original to Koppel et al., “Authorship Attribution in the Wild,” 84.

¹⁸ By our count thirteen largely multivariate stylochronometric studies exist to date. See for example Forsyth et al., “Investigating the Authenticity of the Consolatio,” 375–400; Can and Patton, “Change of Writing Style with Time,” 61–82; HaCohen-Kerner et al., “Identifying Historical Period and Ethnic Origin of Document,” 102–13; Koppel et al., “Computational Methods in Authorship Attribution,” 15.

components such as gender¹⁹ and ethnicity,²⁰ and idiolectic components such as plagiarism and pastiche.²¹ *Implication:* Pursue these same outcomes in NT studies.

7. While NT studies has witnessed a long history of univariate analysis, only 20 multivariate studies of any stripe have been employed in the NT. We are thus still on the computational linguistics shoreline looking across an ocean of possibilities in terms of disentangling sociolect from idiolect in the GNT. However, advances in exploratory multivariate analysis,²² experimental design²³ software automation²⁴ and

¹⁹ By our count seventeen largely multivariate studies exist to date. See for example Thomson and Murachver, "Predicting Gender," 193–208; Koppel et al., "Automatically Categorizing Written Texts," 401–12; Kucukyilmaz et al., "Chat Mining for Gender Prediction," 274–83; Kucukyilmaz et al., "Chat Mining for Gender Prediction," 274–83; Mukherjee and Liu, "Improving Gender Classification of Blog Authors," 207–17; Cheng et al., "Author Gender Identification from Text," 78–88.

²⁰ For initial ethnicity studies see HaCohen-Kerner et al., "Identifying Historical Period and Ethnic Origin of Document," 102–13; HaCohen-Kerner et al., "Stylistic Feature Sets as Classifiers," 847–62.

²¹ By our count twenty-one largely multivariate studies in plagiarism detection have been published. See for example Eissen and Stein, "Intrinsic Plagiarism Detection," 565–69; Barrón-Cedeño et al., "Plagiarism Detection across Distant Language Pairs," 37–45; Stein et al., "Intrinsic Plagiarism Analysis," 63–82; Kuta and Kitowski, "Optimisation of Character N-Gram Profiles," 500–11. For pastiche detection see Sigelman and Jacoby, "The Not-So-Simple Art of Imitation," 11–28; Somers and Tweedie, "Authorship Attribution and Pastiche," 407–29; Dinu et al., "Pastiche Detection Based on Stopword Rankings: Exposing Impersonators of a Romanian Writer," 72–77.

²² See especially the following studies on Principle Components Analysis (Hotelling, "Analysis of a Complex of Statistical Variables," 498–520; Loehlin, "Component Analysis Versus Common Factor Analysis," 29–31; Holmes, "A Stylometric Analysis of Mormon Scripture," 110–14; Husson et al., "Principal Component Analysis (PCA)," 1–58); Factor Analysis (Thurstone, *The Vectors of Mind*, 44–198; Moerk, "An Objective, Statistical Description of Style," 51–58; Kim and Mueller, *Factor Analysis Statistical Methods* passim; Biber, *Variation Across Speech and Writing*, 61–198; Biber, "Linguistic Analyses of Genre Variation," 333–43; Mealand, "The Extent of the Pauline Corpus," 64–90; Argamon and Koppel, "Finding Meaning in Stylistic Variation," 94) and Correspondence Analysis (Carroll, "Vectors of Prose Style," 283–92; Benzecri, "Elaboration statistique de données sur ordinateur," 7–30; Mealand, "Measuring Genre Differences in Mark," 227–45; Clausen, *Applied Correspondence Analysis*, 2–43; Linmans, "Correspondence Analysis of the Synoptic Gospels," 1–13; Mealand, "Style, Genre, and Authorship in Acts," 479–505; Greenacre, *Correspondence Analysis in Practice*, 1–160; Husson et al., "Correspondence Analysis (CA)," 59–126; Mealand, "Is there Stylometric Evidence for Q?," 483–507; Mealand, "Hellenistic Greek and the New Testament," 323–45; Kestemont et al., "Robust Rhymes," 54–76). For an excellent general introduction see Husson et al., *Exploratory Multivariate Analysis*, 1–168.

²³ See especially the following works for fundamental experimental design theory and praxis; Fisher, *The Design of Experiments*, passim; Spector, *Research Designs*, 39–76; Brown and Melamed, *Experimental Design and Analysis*, 20–44; 50–79; Rudman, "The State of Authorship Attribution Studies," 359–61; Gray, *Doing Research in the Real World*, 15–62; Atkinson, *Optimum Experimental Designs, with SAS*, 25–33; 58–71; Meyers et al., *Applied Multivariate Research*; Rasch et al., *Optimal Experimental Design with R*, 175–288.

²⁴ The two most recognized software automation tools are WEKA and JGAAP. Our meta-analysis has identified eighty-three studies that have either used or mention WEKA. For Juola's JGAAP software see Juola et al., "A Prototype for Authorship Attribution Studies," 174–76; Juola and Vescovi, "Empirical Evaluation of Authorship Obfuscation Using JGAAP," 15–17; Juola, "Large-Scale Experiments in Authorship Attribution," 277–82; Juola, "Detecting Stylistic Deception," 92–95; Juola et al., "Towards Active Linguistic Authentication," 386–95.

large-scale experimental design execution²⁵ hold out substantial promise in assisting researchers to disentangle authorship from genre in the GNT. *Implication*: Pursue software automation and large-scale experimental design execution in our experimental design.

8. Two unexplored areas exist in computational stylistics: the use of latent structure or structural equation modelling²⁶ to disentangle those factors, and subsequent causal and directional conclusions that may be drawn from them.²⁷ *Implication*: Pursue latent structural and causal modeling in terms of the NT.

4.2 Implications from the Design Review

The findings summarized in Section 4.1 cover over 250 multivariate studies.

Problematically, however, these studies have been executed upon disparate corpora operating upon different languages from different time periods using different linguistic measures and analyzed with different methods. Thus, while these studies provide a rich starting set of implications (per 4.1) they point to other necessary implications:

²⁵ See especially Park and Zhang, “Large Scale Unstructured Document Classification,” 90–98; Park and Zhang, “Co-trained Support Vector Machines for Large Scale Unstructured Document Classification ” 424–37; Madigan et al., “Author Identification on the Large Scale,” 2–17; Genkin et al., “Large-Scale Bayesian Logistic Regression,” 292–302; Luyckx, *Scalability Issues*, 11–130; Luyckx and Daelemans, “The Effect of Author Set Size and Data Size,” 36–53; Juola, “Large-Scale Experiments in Authorship Attribution,” 275–82; van Dam and Hauff, “Large-Scale Author Verification: Temporal and Topical Influences,” 1039–42.

²⁶ Classic latent covariance structure modeling (SEM, or LISREL) has been only been rarely deployed in computational stylistics. Some tentative studies have been pursued. See Gries, Biber, and Paolillo et al. (Gries, *Multifactorial Analysis in Corpus Linguistics*, 178–80; Biber, *Dimensions of Register Variation*, 140; Biber, *Variation Across Speech and Writing*, 115–20; Biber, “Linguistic Analyses of Genre Variation,” 340; Paolillo et al., “Genre Emergence in Amateur Flash,” 277–301). For nonparametric latent structure work see Dai and Storkey, “The Supervised Hierarchical Dirichlet Process,” *passim*.

²⁷ For an extended discussion of the mathematical determination of directionality see Davis, *The Logic of Causal Order*, 7–68.

1. *Theoretical Implications.* We don't lack ideas *per se*. We lack a theory²⁸ and a method for getting us there. Per Section 2.1 I suggest an abductivist/IBE framework to assist in developing and executing a synthetic experimental design.²⁹
2. *Linguistic Implications.* The epic contradictions in the univariate literature that caused the two "hot messes"³⁰ in computational stylistics were decidedly not chiefly because scholars were incompetent. Language is simply far more complex than we ever imagined. By means of an analogy, we were attempting to take in the Grand Canyon by looking through the straw of traditional grammar via a limited set of *ad hoc* markers. Accordingly the following steps are indicated:
 - a. Use the largest natural language-in-use corpora available.³¹ Our eventual scope must be to embrace all of the Koine, but for pragmatic reasons we must use the largest corpora currently available to us, the mixed-genre texts of the GNT.
 - b. Use language theories amenable to empirical quantification. Because Systemic Functional Linguistics (SFL) provides a more adequate framework than traditional grammar (per Section 2.4), comprehensive SFL measures should be

²⁸ See especially Todorov, "The Place of Style," 29; Rudman, "The State of Authorship Attribution Studies," 361; Juola and Baayen, "Authorship Identification by Cross-Entropy," 64–65; Juola, "Author Attribution," 233; Juola, "Large-Scale Experiments in Authorship Attribution," 275; Argamon, "Scalability Issues in Authorship Attribution," 95.

²⁹ See especially, Peirce, "On the Natural Classification of Arguments," 261–87; Peirce, "Deduction, Induction, and Hypothesis," 470–82; Peirce, "On the Logic of Drawing History from Ancient Documents," 75–114. Concerning the likely unrecognized connection between Peirce and Polanyi, see Mullins, "Peirce's Abduction," 199–216. See also Prendinger and Ishizuka, "A Creative Abduction Approach," 324–26.

³⁰ Rudman, "The State of Authorship Attribution Studies," 351. Cf. also Juola, "Authorship Attribution for Electronic Documents," 121. Regarding the second mess see Argamon, "Scalability Issues in Authorship Attribution," 95.

³¹ Hudson summarizes the notion powerfully by stating that, "to study speech without reference to the society which uses it is to exclude the possibility of finding social explanations for the structures that are used." Hudson, *Sociolinguistics*, 11.

developed, inclusive of the multiple strata,³² syntagmatic structures, paradigmatic systems, and metafunctions of the GNT.³³

- c. Use all language measures when performing unsupervised (exploratory) analysis rather than subsets of measures.³⁴
- d. Use representative (FSE-selected) measures when performing supervised (exploratory) analysis.
- e. Explore the linear information flow of the Koine and its cohesive structures in the discourse structure of the text. We should think like linguists and study language *in vivo* rather than merely pulling it apart and aggregating its eviscerated pieces.³⁵

3. *Experimental Design Implications.*

- a. Build out a formal synthetic experimental design constituted of experiments and subexperiments to achieve these ends. (See Section 4.3.)
- b. Test alternative causes for all significant findings *before* issuing conclusions. This eliminates Type II error.
- c. Implement a broad-based initial exploratory phase of unsupervised experimentation first. This will yield a fundamental understanding of the components and causes of variation.

³² See especially Gliozzo and Strapparava, *Semantic Domains in Computational Linguistics*, 13–48. For now the domains of Louw & Nida (Louw and Nida, eds., *Greek-English Lexicon of the New Testament, Volume 1*) will suffice as a starting point.

³³ Because abduction requires all relevant testable language schema to be tested (per IBE) Dik's functional grammar and Lamb's stratificational grammar will also be tested.

³⁴ Val Haltern makes the same conclusion, "we make no a priori choice for a specific type of word (or more complex feature) to be counted. Instead, all possible features are included" Van Halteren, "Author Verification," 2. See also Gurney and Gurney, "Subsets and Homogeneity," 133.

³⁵ For information flow and cohesion respectively see especially Ghadessy, ed., *Thematic Development in English Texts*, 132–33; Halliday and Hasan, *Cohesion in English*, 1–30.

- d. Once the components at each layer of language are developed and understood, develop and justify more objective methods to map them to their putative causes of idiolect, sociolect, and dialect, etc.

4. *Quantitative Method Implications:*

- a. Compare the components developed by *a posteriori* multivariate methods of extracting authorship and genre analysis (via EMVA/eigen-systems methods) to the main effects and interaction effects developed by *a priori* multivariate analysis (such as log-linear modeling).³⁶ Use a variety of commonly accepted theories of authorship from the history of NT scholarship to feed the *a priori* designs.
- b. In addition to simple latent analysis (factor analysis), implement the “higher octane” forms of structural analysis: structural equation modelling (SEM) and latent structural modeling (LISREL). Use these to help explain both the implicit causality and directionality of the latent components so derived.³⁷ Graphical models are also clearly indicated.

5. *Verificational Implications.* In addition to classical (Fisherian) statistical inference, implement parallel methods³⁸ such as Bayesian inference, bootstrapping and mixed modeling to provide a measure of independent confirmation.³⁹

³⁶ The latter can be decomposed via log-linear modelling.

³⁷ Structural models and graphical models recognize that in complex systems latent, unobserved variables exist between observed effects and a presumptive cause. Graphical models such as Bayesian networks are often used in topic and genre classification. For the capability of component extraction to demonstrate directionality and hence causality see especially Davis, *The Logic of Causal Order*, 7–34.

³⁸ For the initial example of the this kind of work, see Mosteller and Wallace, “Inference in an Authorship Problem,” 282–302.

³⁹ See, respectively, Iversen, *Bayesian Statistical Inference*, 7–17; Mooney and Duval, *Bootstrapping: A Nonparametric Approach*, 1–14; Tashakkori and Teddlie, *Handbook of Mixed Methods*, 285–313.

4.3 Assemble Implications from Current Large-Scale Work

Given the implications we have just assembled, there remains only one more question that concerns us: what implications from *current* experimental design theory should we implement within our synthetic experimental design? Answering this requires that we demystify what “experimental design” actually means to the contemporary experimentalist. Brown and Melamed begin by characterizing experimental design as “an extension of inquisitiveness.”⁴⁰ It is surely that. But it is more; it also concerns itself with reducing the two classic types of errors, false positives and false negatives, and then probing for errors behind those errors.⁴¹ This, indeed, is what most modern researchers think of when they consider experimental design theory or weight different research designs.⁴² Their historic mentor in this, perhaps to no one’s surprise, was R.A. Fisher who penned that experimentation is “experience carefully planned in advance.”⁴³ What Fisher first laid out in 1935,⁴⁴ and subsequently inculcated in generations of researchers even to this day, might be called the central dogma of experimentation. This is simply the notion that the researcher must hold all the variables in a design constant, change them one at a time and observe the results. This surely is an important notion, but it provides little succor if we don’t know what the variables are, or how they may interact to create the things that we see. What we have in language in general is decidedly *not* a simple system of known variables—variables that we can isolate from one another and perturb one at a time. Moreover, in terms of our own particular concern, the Greek of the New Testament, we cannot perturb anything *at all* in a language no longer spoken, and in a corpus unchanged for almost two

⁴⁰ Brown and Melamed, *Experimental Design and Analysis*, 1.

⁴¹ See for instance Viswanathan, *Measurement Error and Research Design*, 97–238.

⁴² See Spector’s classic SAGE monograph (Spector, *Research Designs*, 11–76).

⁴³ Fisher, *The Design of Experiments*, 8.

⁴⁴ This was the second of Fisher’s work that redefined the praxis of statistics. His *Design of Experiments* would progress through eight editions before his death.

millennia. To make matters worse, classical (and I mean so in the Fisherian sense) experimental design theory assumes that we are designing a single experiment. What we are attempting to design here, rather, is thousands of experiments on a large scale, one that our abductive cycle can begin to operate upon.

This way of thinking about solving complex problems—building a family of experiments—is relatively new. As with almost all things new, there tends to be a muddle of terms we must work through. A number of researchers have indeed executed “large-scale” experiments. These include studies by Park et al. (2003, 2004),⁴⁵ and Genkin (2007)⁴⁶ in text classification, studies by Luyckx (2005, 2008, 2010)⁴⁷ and Juola⁴⁸ and others⁴⁹ in authorship attribution. Regardless of the study, however, what these authors mean by large-scale, however, is simply “big.” That is, they are attempting to automatically classify *thousands of texts* or authors automatically and quickly—often for the purposes of text retrieval. In the case of Juola, there is an important nuance, however. When he refers to “large-scale” he means *thousands of methods*—or even millions—rolled against the problem of authorship attribution. When we step back from it a bit, however, we see that although Luyckx, for instance, deals with *many texts* and Juola deals with *many methods*, both are focused on building a better “mousetrap”—a better supervised classification model.

⁴⁵ Park and Zhang, “Co-trained Support Vector Machines for Large Scale Unstructured Document Classification” 421–39; Park and Zhang, “Large Scale Unstructured Document Classification,” 88–99. Importantly, Park et al., are attempting to rewrite the standard supervised playbook by adding linguistic information to the standard approach of executing supervised mathematics upon the now traditional “bag of words.” Cf. to the original work by Stamatatos et al., “Automatic Text Categorization,” 471–95.

⁴⁶ Genkin et al., “Large-Scale Bayesian Logistic Regression,” 291–304.

⁴⁷ Luyckx and Daelemans, “Authorship Attribution and Verification with Many Authors and Limited Data,” 513–20; Luyckx, *Scalability Issues*, passim; Luyckx and Daelemans, “The Effect of Author Set Size and Data Size,” 35–55.

⁴⁸ What Juola is attempting is somewhat closer to our goal. Since his original development of JGAAP (a sort of Swiss army knife for automated authorship attribution) Juola has sought to expand the automation boundary of his discipline. See Juola, “Large-Scale Experiments in Authorship Attribution,” 275–83.

⁴⁹ See Narayanan et al., “On the Feasibility of Internet-Scale Author Identification,” 300–14; van Dam and Hauff, “Large-Scale Author Verification: Temporal and Topical Influences,” 1039–42.

Our work, in contrast, is large-scale in the sense that we seek to generate and test hundreds of thousands of *separate experiments and subexperiments* to yield fundamental insights into causation within a corpus. This is decidedly not a domain of application for supervised classification, although supervised models are clearly indicated in later abductive cycles.⁵⁰ What we seek, rather, requires that we *preserve* the variance and covariance structure of the data so that we may draw out its presumed riches. That is, we need all the data—at least to start with—in order to disentangle the layers.⁵¹ The appropriate technique to begin this kind of work—if we are now wearing our mathematical hats⁵²—is, again, exploratory multivariate analysis (EMVA) but automated using automated knowledge discovery. For our purposes here, given the retrospective look into design implication in Section 4.1 and 4.2 and even these brief comments, we have more than enough background to synthesize the main dimensions of such a design. This we do next.

⁵⁰ Recall that supervised models do not explore the variance and covariance structure necessary in assigning latent causation. Per our comment in section 2.4.3.2.2, supervised models “throw out the linguistic baby with the experimental bathwater.”

⁵¹ This is the work pioneered by Carroll, Blacklith and Moerk, and brought to the attention of computational stylistic researchers by Burrows, Biber, and others (see section 3.9).

⁵² There are other more intuitive ways to begin. See section 4.3.8.

4.4 Develop the Initial Three Dimensions of the Synthetic Experimental Design

Readers will recall that chapter three included twelve separate sections (Sections 2-13) of design implications in addition to the design implications in chapter two. The following table (Table 4.1) collects and summarizes those design implications.⁵³ The columns represent the discrete subsections in chapters two and three and the rows represent the collected design implications themselves.

An important interpretive point should be registered concerning Table 4.1. First, a checkmark in a given cell in the table does *not* mean that that particular aspect of the research design was achieved during that period. Rather it means essentially the opposite—that that era called forth the *need* for that component of the experimental design. Thus, Table 4.1 is, in essence, a prospective look at experimental design implications by learning from the past. As such, it is useful for three purposes. First, it is a tool for helping us visualize the three main dimensions of the synthetic experimental design and their subcomponents. Second, and not coincidentally, it follows mixed-methods research design criteria which underscore exercises such as this.⁵⁴ Third, it hints at the scale, dimensions, and experiments of that design to follow.

⁵³ Table 4.1 is itself a distillation of two more comprehensive tables; Tables A-1 and A-2 in the Appendix.

⁵⁴ See constructs such as this in Tashakkori and Teddlie, *Handbook of Mixed Methods*, 137–65.

The Initial Dimensions of an Experimental Design to Disentangle Authorship & Genre

Code	Criteria (from the chapter subsections)	Chapter 2.X					Chapter 3.X											
		2	3	4	5	6	2	3	4	5	6	7	8	9	10	11	12	13
Measurement Model Dimension																		
<i>Textual Scope</i>																		
MTK	All the extant Koine		✓							✓								✓
MTG	Within the GNT	✓	✓							✓								
MTB	Subsets of Texts between Koine and GNT		✓															
<i>Linguistic Model</i>																		
MLH	Use Halliday's SFL as the primary model	✓	✓							✓								✓
MLL	Use Lamb's SG as a model		✓		✓													✓
MLD	Use Dik's Functional Grammar		✓															✓
MLP	Use Pike's Tagmemics as a model				✓													
MLA	Use SIL (South African) as a model		✓															✓
<i>Linguistic Probes</i>																		
MPS	For each strata of the language	✓	✓		✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
MPP	For each paradigm of the language	✓	✓		✓	✓				✓	✓	✓	✓	✓	✓	✓	✓	✓
MPR	For each rank of the language	✓	✓		✓	✓				✓	✓	✓	✓	✓	✓	✓	✓	✓
MPM	For each (meta) function of the language	✓	✓	✓	✓	✓				✓	✓	✓	✓	✓	✓	✓	✓	✓
MFP	For each discourse structure of the language	✓			✓													✓
<i>Features Selected</i>																		
MFC	Use all token w/in each language layer									✓	✓	✓		✓	✓			
MFV	Use most variant tokens in ea. lang. layer										✓				✓	✓		
MFR	Use representative tokens within a layer			✓							✓				✓	✓		
MFF	Explore low/high-frequency bins				✓			✓			✓			✓	✓	✓	✓	✓
MFL	Explore linear information flow; discourse										✓			✓	✓	✓	✓	✓
Grouping Model Dimension																		
GHA	Test authorship hypotheses vs. rest											✓				✓		
GHG	Test genre hypotheses vs. rest							✓			✓	✓			✓			
GHR	Test register hypotheses vs. rest										✓	✓			✓			
GHD	Test date hypotheses vs. rest							✓			✓	✓			✓			
GHS	Test subject matter hypotheses vs. rest										✓	✓			✓			
GHP	Test audience plurality vs. rest										✓	✓			✓			
GHC	Test context of situation vs. rest								✓		✓				✓			
Analytical Model Dimension																		
<i>Bivariate Description and Inference</i>																		
ABA	Association (χ^2 , Cramér's V)										✓							
ABP	Direction (PRE methods)										✓							
ABI	Information (Entropy, UC)										✓							
<i>Multivariate Description and Inference</i>																		
AMP	Parametric (e.g. MANOVA, MANCOVA)					✓					✓							
AMN	Non Parametric (e.g. HLLA, GLL)	✓	✓	✓		✓					✓							
AMI	Information-Theory-Based					✓					✓							
AMB	Bayesian					✓					✓							
<i>Multivariate Extractive</i>																		
AEP	Parametric (e.g. PCA, FA)	✓	✓	✓		✓			✓		✓			✓				
AEN	Nonparametric (e.g. CAMCA, CATPCA)					✓			✓		✓			✓				
AEM	Mixed Parametric (e.g. ALSOS)					✓			✓		✓			✓				
<i>Multivariate Causal (Latent Structural)</i>																		
ACP	Parametric (e.g. SEM, LISREL)	✓	✓			✓		✓			✓	✓		✓				
ACN	Non Parametric (e.g. LDA)		✓	✓		✓					✓	✓		✓				
ACI	Information-Theory-Based					✓					✓	✓		✓				
ACG	Graphical models	✓	✓			✓					✓	✓		✓				
<i>Multivariate Supervised (SMVA)</i>																		
ASP	Parametric (e.g. MDA)	✓	✓	✓		✓					✓			✓				
ASM	Machine Learning (NN, SVM, etc.)	✓	✓	✓		✓					✓			✓				
ASI	Information-Retrieval-Based			✓		✓					✓			✓				
ASE	Ensemble-Based			✓		✓					✓			✓				
Verificational Model Dimension																		
VFC	Use Fisherian (Classical) Statistics	✓									✓	✓						
VBA	Use Bayesian Inference	✓									✓	✓						
VBO	Use Bootstrapping-Based Inference										✓	✓						
VMM	Use Mixed Methods of Inference	✓	✓			✓					✓	✓						
VCM	Use combinations of the above if possible										✓	✓						
VHB	Use multiple hold-back strategies for SMVA	✓		✓		✓												
VDA	Triangulate findings w/ diachronic approaches	✓																

Table 4.1

To best decode the details of Table 4.1 and how it can help us develop our synthetic experimental design a question and answer format will be adopted in the rest of this section.

4.4.1 What are the Main Dimensions of the Proposed Synthetic Experimental Design?

Inspection of Table 4.1 (reading from top to bottom) reveals that three main dimensions are normative for an experimental design in computational stylistics: 1) the use of linguistic measures (*the measurement dimension*), 2) some optional measure of text grouping (*the grouping dimension is not always required*) and 3) an analytical technique/or inferential test operating upon that data (*the analytical dimension*). In addition, there are two other dimensions which are conceptually just as important as the first three: the verificational dimension and the hermeneutical dimension.⁵⁵ Because the verificational dimension can properly be considered as tests “attached” to the analysis dimension, and the hermeneutical dimension is simply the various perspectival frameworks that interrogate the results of an experiment *a posteriori*, neither will figure in the main exercise of this chapter, building out the individual experiments (the cells) of our experimental design. How might we do such a thing? Per the above, since the main dimensions are *language measure*, *text grouping* and *analysis*, all we need do is to view these dimensions as the sides of an experimental design cube. The total number of experiments in the cube is then the product of the “detail” of each of these three sides. Notice that some of that detail (its top level) is indicated in Table 4.1. Thus the top level of detail in the linguistic dimension are its multiple strata, systems, structures and functions. The question now becomes how much detail—how many total categories—are there in each of these three sides? To get a feeling for that we have produced, an exhibit breaking out more detail for two of those elements, *language system* and *language structure*, is in Figure 4.1 below:

⁵⁵ See the Appendix, Tables A-1 and A-2. for the originating tables.

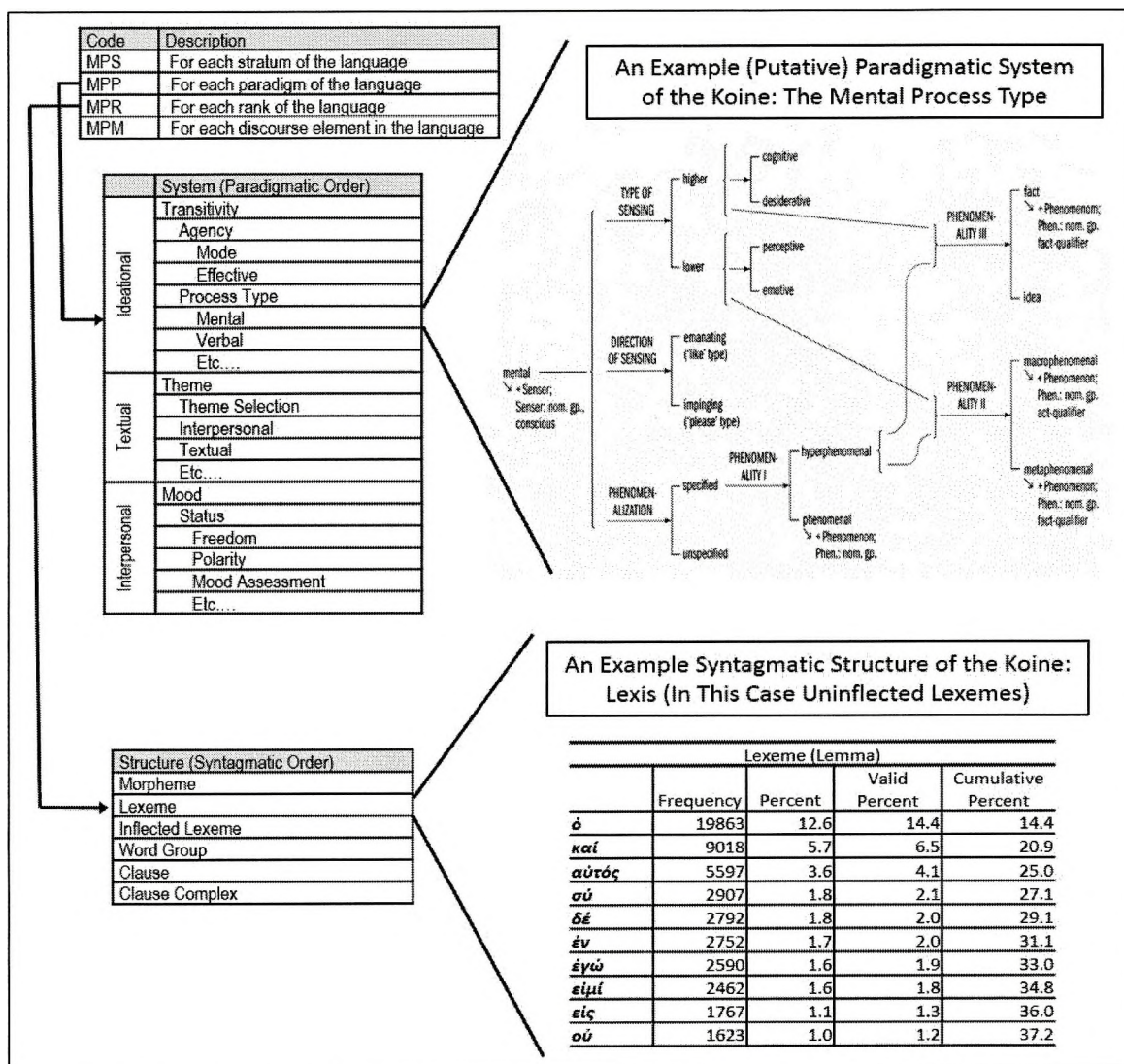


Figure 4.1 An Example of Further Detail (Delicacy) in Structure and System with the GNT

This exhibit breaks out in more detail (delicacy) two main categories underneath the linguistic dimension in Table 4.1. Notice, in particular the category *Linguistic Probes* under the main dimension **Measurement Model**. “Probes” are variables stored in a data repository (in our case the system file of a commercial data analysis package, SPSS) that allow us to rapidly perform computational stylistic analyses. Per Figure 4.1 these probes must be built for each stratum, paradigm (system), rank (structure), and discourse feature of the language. The number of such probes that must be built (its delicacy or the number of its discrete items) is what Figure

4.1 is getting at here. That is, Figure 4.1 “breaks out” one such system and one such structure to a modest level of delicacy. For example, in Figure 4.1 the delicacy level of “fact” within the transitivity system is as follows: IDEATIONAL: TRANSITIVITY: PROCESS TYPE: MENTAL: SENSING: HIGHER: PHENOMENALITY III: FACT. At this particular juncture in the system, then, we are eight levels deep in the system’s *agnation*.⁵⁶ Similar nesting exists in all other systems of paradigmatic choice in the language. It should be clear by now that this exercise is not simply pedagogical—these nested dimensions must fit somehow within an experimental design sequence. This we present in Section 4.4.2 below. Even from this brief discussion it is clear that these many deeply nested linguistic dimensions analyzed by many text combinations and by many types of analyses suggest millions of possible experiments—or more. This in turn requires that some tool or method of automation be created to find all the “legal” combinations of these measures, groupings and methods, generate designs for them, and execute them. The nature of such a tool will be suggested in Section 4.5.

4.4.2 What Is the Typical Sequence Used in Executing a Single Experiment?

By inspecting Table 4.1, reading again from top to bottom, we propose an eight (or optionally nine) step sequence for any experiment in our synthetic experimental design:

1. *Selecting a coherent set of linguistic measures.*
2. *Grouping those measures by hypothesized authorship or genre **text groupings**.*
3. *Analyzing that set of measures with an appropriate analytical technique.⁵⁷*
4. *Testing that set for significance using one or more appropriate inferential tests.*
5. *Interpreting that inference with an appropriate interpretive framework to decide if anything significant was discovered.*
6. *Comparing that significance to other relevant interpretive frameworks (mixed modeling, Bayesian, etc.) applied to the same measures using the same technique.*
7. *Evaluating the results and executing other **subexperiments** as indicated. If indicated, repeat steps 1-5 on those subexperiments before proceeding.*

⁵⁶ Systemic Functional Linguistics, following Halliday, refers to the traversing of a system or locating one’s position within a system as *agnation*. See Halliday, *On Language and Linguistics*, 29. Gleason, apparently, was the first to appropriate the term within anthropology. Gleason, *Linguistics and English Grammar*, 202–08.

⁵⁷ Sometimes that analysis is within the measures itself and sometime it is between text groups.

8. *Storing* the results of steps 1-6 in a data repository for each experiment and subexperiment. Note: The results are stored regardless of significance. This allows a meta-analysis to be performed *a posteriori* to discover if there is a pattern to the significant or near-significant findings.
9. *Returning* to step one and starting over again with a new experiment.

It should be mentioned that while steps 1, 3 through 5, and 7 are quite traditional in computational stylistics, step 2 is virtually absent in studies done to date. Steps 7, 8 and 9 are called for by our abductive design. Each resulting experiment is said to produce one “cell” in the design.

4.4.3 What Else Do We Need to Know in Order to Generate Our Experimental Design?

In order to generate all the experiments and subexperiments in our experimental design, we should take into consideration three more pieces of the puzzle: 1) *its further possible analytical dimensions*, 2) *the integration of the analysis/inferential technique* and 3) *its cell-level validity*.

Its further possible analytical dimensions. This refers to the fact that, though our main analytical dimensions are a given *language measure* (e.g. all the instances of lexis) by a given *text grouping* (e.g. the 15-author theory of GNT authorship) by a given *analysis/inferential theory* (e.g. Cramér’s V), the cells of the experimental design can go quite deeper. That is, we may wish to understand more fully the *interaction* between a given authorial theory (e.g. the 15-author theory) and a given genre theory (e.g. the 12-genre theory of the GNT). This would require a *three-way* contingency table (semantic domains x authorship theory x genre theory) not a two-way table (semantic domains x author theory). Moreover, in this case we would employ a mathematical technique useful for calculating interactions such as log-linear analysis. Thus our design generator must be architected so as to generate more deeply nested experiments. This leads to our second point below.

The integration of the analysis/inferential technique. In the above example the analytical test (Cramér's V) and the inferential test are *integrated*. That is, when the researcher calls upon the commercial software (or the R programming language) to calculate Cramér's V, the software typically performs such inferential work as a matter of course. Not all advanced techniques, however, are so integrated. In the case of log-linear analysis, for instance, many commercial statistical packages do not automatically run all relevant tests of inference on model-based methods such as log-linear analysis. Thus, pragmatically, for some cells (experiments in the design) the generator must be "smart" enough to know this and call for the appropriate test of inference posterior to the analysis if it is not done automatically.

Cell-level validity. This refers to the fact that not all cells (experiments) in the main design are, indeed, valid in an experimental design. As a simple case in point, let's assume the experimental design generator creates the following cell: a two-way contingency table of *linguistic measure* (semantic domains) by *text grouping* (the 15-author theory) analyzed by *analytical technique/inferential test* (t-test). Such a cell is invalid because t-tests assume scaled data rather than the categorical (frequency) data of a contingency table. This raises the issue of experimental design constraints which we discuss below.

4.4.4 How Big Is It? How Many Total Experiments Will Be in the Experimental Design?

Per the discussion in Section 4.4.1, the total number of experiments that *can* be run (which is greater than the number that *should* to be run) would normally be equal to the product of the three main dimensions. This requires that we discover (a) the number of discrete linguistic categories that exist in the Koine, including their combinations, (b) the number of discrete authorship vs. genre combinations that we wish to explore and (c) the number of discrete analytical tests of the same. Recall, per 4.4.3, that only *valid* experiments should be run. Thus, all invalid experiments must be subtracted from this total. In fact, as we will see shortly, the causes

of invalidity in experimental design build-outs can be quite high. To assist in explicating this, the total design build-out can be visualized (as we previously mentioned) as a cube with the three aforementioned dimensions as its axes. See Figure 4.2.

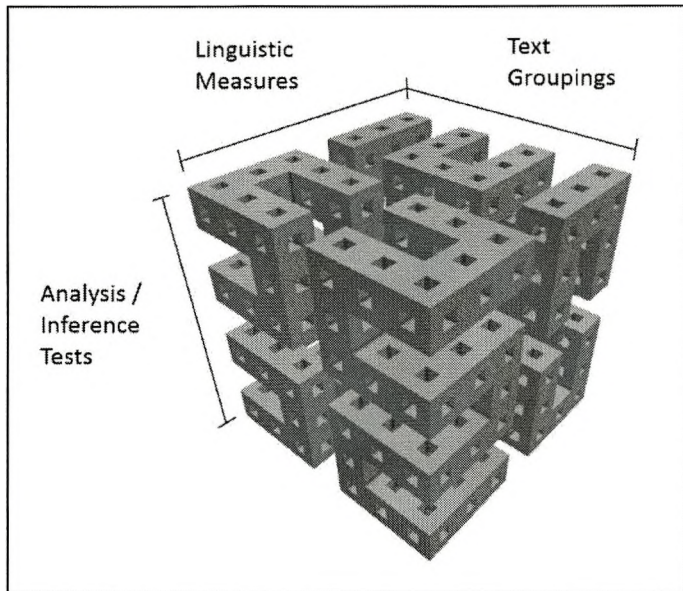


Figure 4.2

number of valid experiments in our synthetic experimental design.

First, in unsupervised experiments we can eliminate *some* of the dimensionality because unsupervised mathematics applied to linguistic data develops text grouping (often sociolectally or idiolectally interpreted components) *a posteriori*. Thus, at that level the design is two-dimensional. In this case we still submit all the discrete linguistic measures, one linguistic class at a time, to these unsupervised approaches (e.g. factor analysis, PCA, CATPCA etc.) and perform the extraction of the resulting components.

Second, per the entire discussion just completed, how many valid cells or individual experiments exist, then, in our experimental design “cube”? To calculate that total number we need to fully break out the detail of each of the three main dimensions. This we address next.

Notice, first, that not every cell in this square⁵⁸ is “solid”; the design has voids in it because not every intersection of these three dimensions yields a valid experiment per the discussion in 4.4.3.

Finally there are two tertiary pieces of the puzzle we must grapple with to get to our estimate of the total

⁵⁸ This notional visualization is of a Hilbert square. The total set of experiments would look visually similar (it would have voids) but it would not be topologically identical to a Hilbert square.

4.4.4.1 Calculating the Viable Combinations for Each of the Three Experimental Design Dimensions

Calculating the Linguistic Dimension

In calculating the linguistic dimensionality of the Koine we will simplify our approach if we limit our exploration to its paradigmatic potential (all the possible categories of choice in the language), the syntagmatic potential (all the possibilities of its “slot and filler” structure) and the extra-lexicogrammatical stratal potential (here we will consider only semantic domains) of the language. Per my now common practice, I will treat each in turn.

First, how would we go about calculating the *paradigmatic potential* of the Koine? To do so requires that we first count the total number of discrete unisystematic *choice-points* across all the systems of the language. Halliday terms the binary oppositions in a given system network (what we term choice-points) as “an alternative with an entry condition.”⁵⁹ Just as importantly for our purposes, Halliday points out that such choices are *not* generally equiprobable. This is actually salutary for our effort, because past a certain level of delicacy we need not generate experiments whose realization in the corpus falls below a given frequency. For the time being we will ignore this fact, however, and assume a corpus of infinite size. This will require “walking down” to the terminal choice-points in all systems in order to calculate all their viable combinations. (We will assume that ordering, *permutations*, are not important at this juncture.) Accordingly, how many unisystematic choice-points are there? Let us assume, as a ballpark estimate, an average of perhaps thirty such choice-points per system,⁶⁰ and that perhaps twenty such systems exist in the system network (i.e. all systems) of the Koine. This would yield conservatively on the order of 600 unisystemic choice-points or possible categories of selection

⁵⁹ Halliday and Matthiessen, *An Introduction to Functional Grammar*, 22–24.

⁶⁰ See Halliday’s modest listing of such systems in English, Halliday, *On Grammar*, 127–51.

expressions. Calculating the total combinations of these 600 choice points is modestly more challenging. The reason for this is that while each choice in a given single system is typically binary and exclusive (though it might be grammaticalized by hundreds of different lexical forms) *multiple semantic systems are simultaneously operative in the grammaticalization of that choice*—systems of transitivity, mood, and theme in particular.⁶¹ That is, at each successive “word-choice” in, say, a five-word clause, that choice exists in the context of other prior choices and is simultaneously expressing different “layers” of meaning as each word is added in the linear outflow we perceive as language. To calculate these combinations we may use any number of strategies. Combinations across the multiple dimensions can be calculated, for instance, using the mathematics of Stirling numbers of the second kind. That type of mathematics need not detain us here, but assuming that the Koine has a system network of the size just indicated, the Koine would yield modestly more than seven million combinations of paradigmatic choice in the GNT.⁶²

Second, what about the syntagmatic cline of the Koine? In terms of one SFL annotation schema related to syntagmatics, the schema of OpenText,⁶³ we know that value explicitly; there

⁶¹ See Halliday’s introduction to this notion in Halliday and Matthiessen, *An Introduction to Functional Grammar*, 58–63.

⁶² This number can be approximated using the simpler mathematics of Pascal’s triangle. Assume, for instance, that 350 of these 600 choice points exist in combinations with other systems, but not within their own system. Then the number of potential combinations with other systems may be approximated as 350 items taken two at a time = 61,075, plus 200 of those taken three at a time = 1,313,400, plus 100 of those taken four at a time = 3,921,225, and 50 of those taken five at a time = 2,118,760. This yields 7,414,460 combinations.

⁶³ Per Porter and Pitts, “New Testament Greek Language and Linguistics in Recent Research,” 234, OpenText is an SFL-based model of annotation applied to the GNT. The model derives from the works of Porter (Porter, “Word Order and Clause Structure in New Testament Greek,” passim; Porter, *Idioms of the Greek New Testament*), Reed (Reed, *A Discourse Analysis of Philippians*) and O’Donnell (O’Donnell, “The Use of Annotated Corpora,”; O’Donnell, *Corpus Linguistics*) and their joint work (O’Donnell et al., “OpenText.org and the Problems and Prospects of Working with Ancient Discourse,” passim; Porter and O’Donnell, “Theoretical Issues for Corpus Linguistics,” passim).

are 49,010 actual (realized) syntagmatic linguistic categories in the GNT.⁶⁴ That number can be known with precision because all we need do is run frequencies on an appropriately annotated GNT (per Section 5.2). If we assume (another simplification) that we limit our exploration to lexemes, inflected lexemes (words), word groups, clauses and clause complexes, the frequencies of those forms that occur at least once in the GNT is 49,010.⁶⁵ Yet this number relates to the NA27 text of the GNT. What we are interested in is a *corpus-invariant* experimental design. That is, our design depth (albeit not our first abductive cycle) must be directed to the larger Koine not just the GNT. Moreover, this does not address combinations *at the level of discourse*. Here we must add, at the very least, high-frequency colligations found within spans of say, one to ten words from each other. Let's assume then, another two to three million possibilities as a rough estimate of the paradigmatic potential of the Koine.

Third, using the semantic domain definition of Louw and Nida,⁶⁶ 756 semantic minor and major domains are extant in the GNT (see Section 5.2). Ignoring other extra-lexicogrammatical measures other than semantics,⁶⁷ we will assume the upper limit of semantic subdomains in the Koine to be, say, 2,000. This completes our modest exploration of the linguistic potential of the Koine. The total number of "items" that must be tested on the linguistic axis of our cube is therefore the sum of the paradigmatic, syntagmatic, and semantics categories—and it is a large number indeed, more than 10 million categories. Now we turn to consider the same exercise in terms of the other two sides of our design cube.

⁶⁴ This value was arrived at by summing the actual frequencies of the lexemes, words, word groups, clauses and the two types of clause complexes. There is overlap in the categories of the two types of clause complexes so this still leaves us with over 39,000 grammaticalized categories of syntagmatic structure in the GNT.

⁶⁵ This justifies, it should be mentioned, the faith placed in the paradigmatic and syntagmatic richness of SFL argued for in section 2.4.

⁶⁶ Louw and Nida, eds., *Greek-English Lexicon of the New Testament, Volume 1*.

⁶⁷ E.g. phonemic and morphemic levels of strata.

4.4.4.2 Calculating the Analytical and Text Grouping Dimensions

The analytical/inferential and text grouping dimensions are substantially easier to calculate. Three important points concern us here as well. All three points make a difference on how we “count” the available methods. First, while the analytical dimensions themselves may also be described in terms of binary oppositions or trinary oppositions (e.g. univariate vs. multivariate methods, parametric vs. nonparametric methods, etc.) the relationship here is taxonomic; a given method is classed in multiple levels of the taxonomy simultaneously. For example, factor analysis is a parametric exploratory multivariate method capable of being used for structural studies of both causality and directionality. Accordingly, we must be careful to avoid “double counting” factor analysis more than once. Second, many of the methods that were historically prevalent in the history of computational stylistics have been largely superseded by other methods (e.g. chi-square). Third, many of the more advanced multivariate methods exist in multiple forms. That is, their objective function (what they are tasked with doing) can be arrived at via multiple underlying algorithms. (Thus, discriminant analysis, for instance, may be performed in a stepwise fashion by either Rao’s V or Wilk’s Λ .) Even by factoring in these complexities, it is likely safe to say that adding up all the univariate, multivariate and other mathematical modeling methods (including supervised and unsupervised mathematics, but not their combinations or ensembles, will yield on the order of around 1,000 different mathematical methods useful for computational stylistic analysis. Lastly, we will assume that there are only a handful of viable authorship and genre candidate theories (say ten).

Summing the Potential Number of Experiments

Given that there are approximately ten million linguistic species on one side of our experimental design cube by over 1,000 mathematical methods on the second side of the cube, by at least ten text groupings (here we ignore the text grouping combinations) on the third side,

this yields on the order of 100 billion potential experiments. Given that many of these experiments will be flagged because of invalidity (per Section 4.4.3) it may be safe to estimate that the total number of experiments we will need to run is on the order of ten billion.

4.5 Explore Technologies to Automate the Execution of the Experimental Design

4.5.1 One Proposed Component: Software Automation

The conclusion to the last section may boggle the mind. If there are on the order of ten billion potential experiments on the Koine, how can such a thing ever be achieved? The answer is simple. It cannot—at least not without thinking both “harder” and “smarter.” Even if by thinking harder however (such as eliminating illegal experiments and pruning the paradigmatic richness of our exploration), we still need to assay a representative percentage of the Koine—one that will still allow us to capture the rhythm of the language of the Koine in all its quantitative richness. Let’s assume that we could do so by reducing the total number of experiments 100 fold. This would still yield something like 100 million experiments. If so, this clearly requires, per Section 4.4.1, some “tool” that can (a) find all the legal combinations of measures, groupings and methods, (b) generate designs for them, (c) execute them—preferably with some high level of automation, d) evaluate some large number of them, and lastly (e) submit the rest to us—with a report of its own findings, of course. For the time being we will define this tool as having two components: an experimental design generation module that generates all the legal experiments (the generator), and an experimental design execution module that executes these experiments (the distributed computing engine). These notions will be further detailed in Sections 4.4.3 – 4.4.9.

4.5.2 What are Experimental Design Constraints and Why Do We Need Them?

Notice the underlined adjectives in Section 4.4.2. Almost all experimentation requires that a set of prerequisites be met to avoid various sources of error. That is, experiments must

meet simple (and sometimes not so simple) sets of requirements. Those requirements are also termed experimental *constraints*. Per Section 4.3.3 above, as far as the generator is concerned, these constraints are relatively simple logic statements, things that check and catch the kind of inconsistencies described above. If any of the requirements are not met for any given cell for any reason, the generator will not instantiate (write out) that cell for execution by the experimental design execution module. Five general classes of requirements/constraints exist: *analytical fitness* (analysis category appropriateness), *distributional adequacy* (the data fits distributional prerequisites for the test), *degrees of freedom adequacy* (the sample size of the data is large enough to yield valid results), *theoretical salience* (whether the test explores a salient task)⁶⁸ and *pragmatic currency* (e.g. the task is “doable” right now). As an example, if our GNT has not yet been annotated with a target language schema such as, say, Lamb’s stratificational grammar annotation, then clearly no stratificational grammar-based hypotheses can be executed.

4.5.3 How Can Such a Massive Experimental Design be Generated?

Despite the massive number of cells in such an experimental design, for researchers in the field of knowledge discovery the scale of this task is, in actuality, only a modest knowledge discovery effort. Especially in the worlds of commercial data mining (business intelligence) and bio-informatics, automated knowledge discovery (AKD) is an active area of research and practice.⁶⁹ Specifically, software code generators – software that writes commands which are

⁶⁸ For example, theoretical salience would interdict building a model that compares the interaction of two theories that are too closely specified. Not only would such a notion be theoretically uninteresting, it is likely to violate the independence assumptions of some multivariate tests.

⁶⁹ Automated knowledge discovery (AKD) was birthed in the 1970’s as an outgrowth of artificial intelligence (AI) research. It first explored whether computers could develop hypotheses themselves (Hájek and Havránek, *Mechanizing hypothesis formation*) then explored the hardware and software pragmatics of doing so. While often AKD simply means supervised machine learning (Hsu and Chen, “Subspace Clustering of High Dimensional Data,” 31) the principled practice of harnessing software and hardware assets for AKD is an aggressively growing area of research and practice. See especially Botha et al., *Coping with Continuous Change in the Business Environment*, 223–38; Zytow and Zembowicz, “Mining Patterns at Each Scale in Massive Data,” 139–48. Perhaps the greatest commonly recognized yields in AKD have been the successes in bio-informatics, such as the human genome project and continuing advances in automated genome sequencing.

then faithfully executed by databases and statistical software are no longer new. In the late 1990s through the mid-2000s my own firm, for instance, wrote code generators of similar scale to solve large-scale data mining problems in the financial services industry. These code generators typically generate all the legal possibilities of experiments, and then allow the analytical software to reject experiments with insufficient *degrees of freedom* (a large enough sample size for that experiment to execute the type of statistics in which we are interested). Moreover, my firm has also developed distributed computing-based software with multiple mathematical modeling tools.⁷⁰ This software has been optimized so that, as of December 2014, these tools generate, execute and test over ten million multivariate models per week on a modern laptop computer. Moreover, if the process was wholly automated, with a modest investment the computational time for such a task would be about two months.⁷¹

4.5.4 How Many of These Experiments Need We Run to Get to Our Answer?

With all this talk about computers and software and automation we really need to ask a deeper question. *Do we really need to execute over 10 billion experiments? Or even 100 million?* How many are needed to get an accurate feel for the rhythms and patterns of the Koine in terms of authorship and genre? How much data is sufficient to arrive at a conclusion that eliminates the traditional sources of error? These are, of course, philosophical as well as scientific questions. One approach to answering this question is simply to get on with it; start the process, and begin to find those patterns using modern knowledge discovery (KD) methodologies. A second approach would use a variant of brute force KD methods to prune the synthetic experimental design down to size. Either way, KD is a blatant operational implementation of Inference to the

⁷⁰ The commercial application is called MiningSolve.™ See <http://www.decisionsupportsciences.com/miningsolve.html>.

⁷¹ About \$50,000 USD of computer servers housed within a relatively modest 4U rack in a server cabinet.

Best Explanation (IBE). That is, KD is, in essence, supervised math *applied to our experimental design matrix*. That is, it starts randomly and then converges into coherent volumes within our experimental design “cube” that are most fruitful, and then it spends its time there. In the second approach we use intuition (yes, intuition!) to see how *that* works in getting our task down to size. This will be discussed next.

4.6 Given This Experimental Design—What Should We Chose to Do First?

Hypotheses Drawn from the History of NT Research: Even though our quantitative impulse (per implication 4c in Section 4.1.2) would have us start with EMVA modeling and SEM modeling, as a work in NT studies, we will instead start with the hypotheses delivered over to us in the long history of the discourse on authorship issues in the GNT (See code VDA in Table 4.1). This translates, first, to implementing a phase in which we identify from the history of scholarship six primary theories of authorship and four genre theories respectively (see codes GHA and GHR from Table 4.1). The development of these hypothesis is described in Chapter Five.

Bivariate Analysis. Next, and quite intriguingly, *despite the long history of univariate and multivariate research in authorship studies, no simple bivariate research—research that tests the bivariate association of the central authorship and genre hypotheses against the various linguistic measures in the GNT—has been performed.* This is quite surprising given that testing association by bivariate analysis—simple cross tabulation—is considered one of the most primordial steps in exploratory data analysis. The bivariate analysis in this study will begin by using adjusted chi-squared based methods (Cramer’s V), and proportional reduction of error (Goodman & Kruskal’s tau) methods. (See codes ABA from Table 4.1).

Multivariate Analyses. To the bivariate analyses we will add a variety of information theoretic measures (uncertainty coefficient, entropy, noise, etc. per code ABI). In the following

multivariate phases, these same authorship and genre groupings will be submitted to decompositional analysis via Hierarchical Log Linear Analysis (HLLA) and multivariate visualization and decomposition using EMVA. More on the complete measures, groupings, and analytical choices chosen for the first abductive cycle through the design will be covered next.

5 Defining Testable Language Measures, Authorship Theories and Genre Theories

5.1 Framing the First Cycle Through the Abductive Design (Chapters 6–9)

The wish list we originally developed in Chapter One can now be seen to be the product of the review and experimental design work in the first four chapters. This “wish list”:

- Addresses pertinent issues in the GNT—genre and authorship.
- Uses explicitly linguistically-informed measures that both span the syntagmatic rank structure of the GNT and tests a stratum of language (semantics) beyond lexicogrammar.¹
- Operates upon whole vectors of that linguistic data (all types) first, followed by FSE (feature set extraction) based on the standardized adjusted residual.²
- Tests multiple categories of genre (six) and authorship (four) developed from the literature of New Testament studies and literary theory, thereby taking seriously the long history of diachronically developed views on authorship in NT studies.³
- Tests as a control three separate traditional grammatical (part-of-speech) annotations of the GNT.⁴
- Operates progressively from simple bivariate analysis to multivariate analysis.
- Uses multiple methods of bivariate and multivariate methods to provide multiple inferential “looks” into the data.⁵
- Uses automated software tools to generate thousands of separate model runs to discover an optimal set of models depending on convergent or iterative criteria.⁶
- Operates at the text (book) level not the sample level.⁷

¹ Thus meeting the criteria of codes MPR and MPS in Table 4.1.

² Thus meeting the criteria of codes MFC and MFR in Table 4.1.

³ Thus meeting the criteria of codes GHA and GHG and VDA in Table 4.1.

⁴ Thus meeting the criteria of code AEO in Table A-2 in the Appendix.

⁵ Thus meeting the criteria of code AEM in Table A-2 in the Appendix.

⁶ Thus meeting the criteria of code APS in Table A-2 in the Appendix. This particular software tool was originally written in the C programming language by the author and has since been significantly updated and ported to C++ by a team of developers over the last twenty years at Market Advantage Software Inc. (dba Decision Support Sciences).

⁷ In the first cycle through the experimental design the various quantitative approaches will operate at the text (book) level rather than breaking the text into smaller equal-sized samples. This confers a number of important advantages especially for an inaugural abductive cycle: First, it reduces the degrees of freedom of all analyses with concomitant advantages in avoiding problems of statistical inference that arise with sparse contingency tables. Second it is hermeneutically prior, it coheres with how the texts exist in the Canon. To overcome the fact that column and row totals are not equal, three strategies will be employed: (a) to adjust for the unequal column and row totals we will implement statistical methods designed for such (e.g. Cramer’s V for bivariate analysis or EMVA methods that are standardized as a matter of course such as Correspondence Analysis), (b) when required to compare tables across substantially different degrees of freedom (which will be infrequently) we will use various strategies of binning to reduce the asymmetry of the tables and (c) FSE-based methods predicated on cell-level standardized residuals will be employed in order to select the fewest number of rows (linguistic measures) that yield the maximal variance across the columns (texts in the NT).

In order to most efficiently deliver on this “wish list,” the execution of the experimental design has been divided into five chapters (chapters five through nine). The tasks to be performed (per Table 2.1) are summarized below:

- In Chapter Five I will develop and articulate the theory for the linguistic measures, as well as the genre and authorship hypotheses we will test.
- In Chapter Six I will begin the quantitative analysis by testing the authorship and genre hypotheses using various nonparametric measures of bivariate association.
- In Chapter Seven I will test those same authorship and genre hypotheses using nonparametric multivariate analyses.
- In Chapter Eight I will visualize genre and authorship via EMVA methods.
- In chapter nine we will test genre and authorship via a variety of information theory-based measures
- In Chapter Ten I will make conclusions, reflect on the findings hermeneutically and suggest the next cycle(s) through the abductive loop.

In the rest of this chapter, then, the outline is as follows:

Develop a Testable Theory of NT Authorship and Genre:

- 5.2 Develop and justify a *representative set of language measures* which respect modern linguistic and corpus linguistic theory. Use traditional grammar measures as a control.
- 5.3 Develop and justify a set of *authorship theories* which subsume the major authorship views in the history of New Testament criticism. Inform the authorship theories by experimental design criteria to reduce statistical error.
- 5.3 Develop and justify a set of *genre theories* which respect historical and recent findings in literary, rhetorical and genre criticism as well as recent findings in linguistics and socio-linguistics. Use experimental design criteria to reduce statistical error.

A final housekeeping item is helpful before we begin. Since virtually all analyses in this study were based on contingency (crosstab) tables, for those unfamiliar with the notion a sample table of the top thirty lemmas by selected NT books is included below.

	Matt	Mark	Luke	John	Acts	Rom	...	1 Pet	2 Pet	1 Jn	2 Jn	3 Jn	Jude	Rev	Total
<i>ὁ</i>	2788	1514	2646	2186	2709	1105	...	200	123	359	33	29	55	1889	19863
<i>καί</i>	1178	1091	1469	828	1110	276	...	71	63	132	16	11	21	1123	9018
<i>αὐτός</i>	922	760	1086	769	703	158	...	34	29	102	7	4	7	441	5597
<i>σύ</i>	457	164	446	411	263	130	...	53	21	34	8	10	12	80	2907
<i>δέ</i>	494	163	542	213	554	148	...	28	21	11	0	2	13	7	2792
<i>ἐν</i>	293	135	361	226	279	173	...	50	43	79	8	3	8	158	2752
<i>ἐγώ</i>	261	130	282	518	310	148	...	6	20	57	6	6	7	98	2590
<i>εἰμί</i>	289	192	361	445	278	113	...	13	13	99	6	2	4	110	2462
<i>εἰς</i>	218	168	226	187	302	119	...	42	11	9	3	1	6	80	1767
<i>οὐ</i>	202	118	174	282	111	122	...	13	12	48	5	4	2	67	1623
<i>ος</i>	125	88	190	158	225	90	...	31	19	30	3	5	6	70	1407
<i>οὗτος</i>	147	79	229	239	236	52	...	11	22	39	5	4	9	49	1387
<i>λέγω</i>	293	204	219	269	102	34	...	0	1	4	2	0	2	94	1329
<i>θεός</i>	51	49	122	83	167	153	...	39	7	62	2	3	4	96	1317
<i>στι</i>	140	102	174	271	123	56	...	16	5	76	2	1	4	64	1296
<i>πᾶς</i>	129	68	158	65	171	70	...	18	7	27	2	2	8	59	1243
<i>μή</i>	128	77	140	117	64	80	...	14	6	20	5	2	3	50	1042
<i>γάρ</i>	124	66	97	64	80	144	...	10	15	3	1	2	1	16	1041
<i>εἶπον</i>	212	86	314	211	132	14	...	0	0	4	0	0	1	11	1024
<i>Ἰησοῦς</i>	152	82	88	244	69	36	...	10	9	12	2	0	6	14	917
<i>ἐκ</i>	82	67	87	165	84	60	...	8	5	34	1	2	2	135	914
<i>ἐπί</i>	122	72	161	36	169	31	...	9	3	1	0	1	1	144	890
<i>κύριος</i>	80	18	104	52	107	43	...	8	14	0	0	0	7	23	717
<i>ἔχω</i>	74	70	77	87	44	25	...	5	5	28	4	2	2	100	708
<i>πρός</i>	42	65	166	102	133	17	...	3	2	8	3	1	0	8	700

Table 5.1

5.2 Developing Testable Language Measures

While the eventual experimental design goal of this work is to test all extant language measures (all tokens) in the Koine, the objective in this first cycle is to develop a *theoretically grounded* and *representative* set of measures to use as probes in understanding authorship and genre in the GNT. But which of the various candidate language theories is the most theoretically grounded? Which is the most representative? In terms of *theoretical grounding*, recall from Section 2.5.4 that SFL has been selected as our linguistic schema of choice due to its rich multistructural, multisystemic, and multifunctional nature. In short, it provides us with many more “things to test” than traditional grammar.⁸ In terms of *representativeness* three methods of FSE will be employed: the first by frequency (similar to the “Burrows method”) and the second and third by measures that maximally capture the variance between texts in the GNT.⁹

With this framework now in place, I propose an initial exploratory set of eleven representative language measures. Three are from traditional grammar and serve as controls,¹⁰ and seven are situated at five separate “points” across the rank scale of the Koine. These eleven measures can be classified into three linguistic categories: paradigmatics, syntagmatics, and semantics as follows:

1. *Paradigmatic Measures*: Because no paradigmatic system network of the Koine yet exists, as a proxy (with all its attendant limitations) we will test three separate traditional morphological annotations of the GNT (GramCord, AGNT and Logos).

⁸ Because abduction requires all relevant testable language schema to eventually be tested (per IBE), Dik’s functional grammar and Lamb’s stratificational grammar will also be tested, but this development will await subsequent experimental design cycles. This meets the criteria of codes MLL-MLA in Table 4.1.

⁹ These measures are AAVASR and APASR respectively. For more on AAVASR and APASR see Chapter Eight.

¹⁰ These grammatical categories are derived from three independent grammatical annotations of the GNT: the AGNT, GramCord, and Logos respectively. So that these GNTs can be directly compared grammatically we simplified their inflectional categories to a common core set of 1,635 categories (e.g. a finite verb has one of 540 possible paradigmatic forms: 6 tenses by 3 voices by 3 genders by 2 numbers by 5 cases) such that all words received a categorization from each GNT. In this way any differences seen between the GNTs are necessarily due to principled disagreements in formal inflectional categories (parsing) rather than their particularistic annotation schemas which tend to borrow information from other levels of rank.

2. *Semantic Measures*: Two are from semantics (specifically, the nondisambiguated Louw-Nida major domains obtained from Opentext.org and disambiguated Louw-Nida subdomains from Logos).¹¹
3. *Syntagmatic Measures*: Seven will be collected from five distinct “lengths” in the GNT. The benefits of the syntagms so selected are that they:
 - Span from “small to large” GNT structures (from “words” to clause complexes).
 - Recapitulate intuitive and well-understood categories familiar to traditional language pedagogy (lemmas, inflected lexemes ≈ words,¹² word groups ≈ phrases, clauses, sequences of clauses/clause complexes ≈ sentences).¹³
 - Contribute, via the clause complex measures,¹⁴ to the recent and appropriate interest in linguistics above the rank of the clause.¹⁵

Last, all measures are complete within a category (per the last column in Table 5.2) in contrast to a handpicked subset. At its core using all measures is descriptivist¹⁶ and inferentialist in philosophy, seeking to understand total (descriptivist) and typical (inferential) language

¹¹ Semantics is considered to be a distinct stratum from lexicogrammar. See Halliday and Matthiessen, *An Introduction to Functional Grammar*, 25.

¹² By means of definition, lemmas are the “dictionary form” of words. We treat them separately from words in the analysis because doing so (a) broadens the analysis by differentiating it in terms of uninflected vs. inflected vocabulary and (b) meets criteria MPC. See Table A-1 in the Appendix for this and other measurement criteria requirements.

¹³ The congruence symbol used here is intentional. The correspondence is one of rank, not of function.

¹⁴ The clause complex annotations we have developed are a simple extension to the systemic functional annotation of the clause defined in the OpenText GNT. The annotation starts from the first function slot in the first clause and proceeds sequentially to any clauses linked to the initial clause. An example annotation would be: [IA-2PA-3CcjSP-4cjP]. In this example, the first clause constituent is comprised of a lone adjunct, the second clause in the complex is comprised of a predicator followed by an adjunct, the third by a complement, conjunction, subject and predicator and so forth. Two types of clause complexes were developed. The first type used internal OpenText “pointers” to identify which clauses were “chained” to other clauses. These were termed Standard Attribution Clause Complexes. After executing initial analytics, though, it was discovered that this chaining often omitted prior clause referents to which an existing clause referred. Accordingly, we developed a recursion program to “walk back up” the chain and fill in missing clauses; these we termed Proximity Attribution Clause Complexes.

¹⁵ Specifically, text linguistics (discourse analysis) focuses on structure above the rank of the clause. For applied discourse approaches executed upon the GNT, see Porter and Reed, eds., *Discourse Analysis and the New Testament*, passim.

¹⁶ The label “descriptivist,” for many linguists at least, is the kiss of death. In the view of many this was the iceberg that sank the Titanic of linguistic structuralism. Cf. O’Donnell, *Corpus Linguistics*, 16. But given the perspective of time and distance, one wonders whether certain aspects of the programme of the descriptivists, especially the work of the post-Bloomfieldians, was right in many ways, including its deep desire to understand language *in toto*, in terms of its distributions, rather than reductionistic partitions of it. Black, in Svartvik, ed., *Directions in Corpus Linguistics*, 9, notes that “corpora provides [sic] the possibility of total language accountability.” This was the dream of the descriptivists. Where their program demonstrably failed, though, was in both its positivism and in its technical capacity to achieve its ends.

variation across all instances. A summary of further definitional, developmental, and distributional information regarding these eleven measures can be inspected in Table 5.2 below:

<i>Language Measure Used</i>	<i>Linguistic Definition</i>	<i>What it Measures</i>	<i>How Was It Developed?</i>	<i>Categories/ Instances (NA27)</i>
Lemmas	The canonical form of the lexeme ¹⁷	The breadth of lemmas used (lemma vocabulary)	By arbitrating lemma disagreements between the five GNTs used.	5,413 / 138,019 (All)
Unique Lexeme Form: Words ¹⁸	The distinct forms a lexeme can take in a language	The breadth of lexemes used (lexeme vocabulary)	By arbitrating spelling and diacritical differences among the five GNTs	17,736 / 138,019 (All)
Semantic Domains (Dis-ambiguated)	The sense given an arbitrary sign ¹⁹	Meaning (as defined by Louw and Nida's Major Semantic Domains) ²⁰	By using the disambiguated semantic domains of the Logos Syntactic GNT	93 / 138,019 (All)
Semantic Sub-domains (Dis-ambiguated)	The subcategories of sense given an arbitrary sign	Meaning (as defined by Louw and Nida's Semantic Subdomains)	By using the disambiguated semantic subdomains of the Logos GNT	666 possible, 663 actual / 138,019 (all)
Traditional Grammar: AGNT	None <i>per se</i> ²¹	Parts of speech and their subcategories from the AGNT (Friberg) ²²	By mapping AGNT's TG categories into 1,635 common TG categories ²³	1,635 possible, 633 actual / 138,019 (all)
Traditional Grammar: GramCord	None <i>per se</i>	Parts of speech and their subcategories from GramCord (Boyer) ²⁴	By mapping GC's TG categories into 1,635 common TG categories	1,635 possible, 583 actual / 138,019 (all)
Traditional Grammar: Logos	None <i>per se</i>	Parts of speech and their subcategories from Logos' GNT.	By mapping Logos' TG categories into 1,635 common TG categories	1,635 possible, 617 actual / 138,019 (all)
Word Groups	A nominal with optional modifiers (definers, etc.)	Natural groups of words with clauses (equal to a phrase)	Imported from the OpenText GNT ²⁵	886 actual / 89,679 (all)
Clauses	The core syntax unit in language	Natural groups of clauses	Imported from the OpenText GNT	1,412 actual / 30,919 (all)
Clause Complexes (Two types)	Via working back from any clause to its head clause	Structure above the clause	Derived from the SFL clause annotation model of the OpenText GNT	11,260 (Ppox.); 12,303 (std.) methods / NA

Table 5.2

¹⁷ Lexemes are a basic unit of linguistics, a "family group" of words that share the same canonical form, or lemma.

¹⁸ The construct that traditional grammar labels as a "word," linguists view as an instance of a lexeme. It is a discrete sign comprised of one or more morphemes which, in the GNT, may or may not be inflected.

¹⁹ Signs are the arbitrary entities to which we attach meaning. For a helpful nontechnical explanation, see Roy Harris's translator's introduction to Saussure's *Cours*, de Saussure et al., *Course in General Linguistics*, xi–xii.

²⁰ For background theory on semantic domains see Louw and Nida, eds., *Greek-English Lexicon of the New Testament, Volume 1*, vi–xx.

²¹ Traditional grammar predates linguistics proper, and therefore has no particular linguistic definition.

²² For the annotation theory behind the AGNT, see Friberg et al., *Analytical Greek New Testament*, 797–854.

²³ This mapping from the original categories of each GNT to 1,635 common categories was based on formal rather than functional linguistic principles. This allowed us to compare the grammatical annotations directly. We have reserved the functional versions of each grammatical annotation for later testing.

²⁴ The term "GramCord" is most commonly recognized as the Windows-based search software and/or the Institute founded by D.A Carson and Paul Miller that bears the same name. Both are based on the annotation of the GNT originally developed by James Boyer (Boyer, "Project Gramcord: A Report," 97–99).

²⁵ OpenText has applied a "functional and relational dependency model of syntax to the word group and clause structure of the GNT" per Porter and Pitts, "New Testament Greek Language and Linguistics in Recent Research," 234. The model has been in development since the mid-1990s. The OpenText GNT may be viewed broadly as an implementation of SFL principles applied to the Koine. Its clause annotation model is articulated at <http://www.opentext.org/model/introduction.html>.

5.3 Developing Testable Authorship Theories

Our task in Sections 5.3 and 5.4 respectively, is to develop *testable* authorship theories (specific aggregates of texts for a given authorship theory) and *testable* genre theories (specific aggregates of texts for a given genre theory). To develop testable theories, two points must be clearly understood. *First*, regardless of whether one is a biblical scholar, literary scholar, statistician or a linguist, *testable* theories are simply different proposed groupings of the 27 NT texts. *Second*, developing the most defensible and representative sets of authorship and genre theories requires, in my view, that we interrogate all four disciplines just listed. Accordingly, in Section 5.4, all four disciplines are relevant and will be used to develop the genre categorizations. Here in Section 5.3, however, two disciplines in particular, statistical design and the NT scholarship have the most direct bearing on developing testable authorship theories. We therefore start by exploring principles from both statistical theory and biblical scholarship to develop the inaugural set of authorship theories that will be tested in this work.

5.3.1 Principles from Statistical Sampling Theory (Which Also Apply to Genre)

The goal of this section is to develop discrete testable authorship categories in the widest possible *currently testable* compass. That is, while we would certainly prefer to situate the texts of the GNT in terms of the Koine, the widest possible *currently testable* compass of our analysis is the GNT, and not the larger Koine. The reason for this is pragmatic rather than theoretical—six out of ten of the linguistic measures in Section 5.2 currently exist only in terms of the GNT.²⁶ Given that the complete GNT is now our scope, the next question may be “Why not use a smaller subcorpus of the GNT to achieve the disentangling of genre from authorship?” The use

²⁶ To be clear, the OpenText.org text annotations are being added Koine Greek texts contemporaneous to the GNT, but that work is ongoing and nothing remotely resembling a linguistic annotation of known Koine texts contemporaneous to the GNT currently exists.

of a smaller subcorpus (a Pauline sub-corpus) is, in fact, the *de facto* approach used in NT studies. In response, two logical and two experimental design reasons exist that contraindicate starting with a smaller “only Pauline” or “only Johannine” corpus. First, starting only with texts a given scholar considers to be Pauline (or pseudo-Pauline) quite clearly begs the question. If we limit our exploration only to the four, seven or thirteen texts variously ascribed to Paul, for instance, any number we choose may eliminate texts which may be more linguistically Pauline than the number we have chosen. Second, a smaller corpus may also eliminate one or more genres that exist in the larger context of the Koine. If no examples of that NT genre exist within the comparison set of the larger Koine, that given NT text would necessarily and quite erroneously be classified either in the next closest category or as a *sui generis*.²⁷ Third, from an experimental design stance, only when we situate the variance structure of the various Paulines within the larger variance structure of the Koine will we gain enough relief to propose what *is* and what *is not* Pauline.²⁸ Fourth, a smaller corpus increases statistical error.²⁹ Given these four factors, it would clearly be best to run our analytics on all extant literary (as opposed to documentary) texts of the Koine, or at least those contemporary to the New Testament.³⁰ Because, however, such an annotated corpus is currently unavailable, our analyses will be

²⁷ Relevant data not captured inside a statistical model is termed exogenous data. The presence of substantial exogenous variables increases error in prediction and reduces the fit of causal models. For a seminal articulation of this see McQueen, “The Nature of Causality,” 42–43, who not only describes this but deftly situates it within the history of science.

²⁸ “Variance structure” refers to the various ways of understanding the variation or covariation between multiple variables. For an accessible explanation of correlation, and covariance in factor analysis (which simplifies the covariation matrix) see Green and Tull, *Research for Marketing Decisions*, 418–39. For a more mathematical treatment see Kim and Mueller, *Introduction to Factor Analysis*, 12–21.

²⁹ In statistics, the larger the number of observations the smaller the uncertainty. The classical statement of this was articulated in Fisher’s 1936 address to the Harvard Tercentenary Conference of Arts and Sciences. “Clearly there can be no operation properly termed estimation until the parameter to be estimated has been well defined, and this requires that the mathematical form of the distribution shall be given.” Fisher, “Uncertain Inference,” 405.

³⁰ Cf. also O’Donnell, *Corpus Linguistics*, 137.

executed on the next largest set of texts for which we do have modestly complete language measures – the Greek of the New Testament.

5.3.2 Principles from Biblical Scholarship

We propose that two principles govern our efforts in developing representative authorship theories. The first needs no defense, namely, that we test all relevant historic views of Paul, not just current construals of him. The second is like it: test all reasonable variants of current construals of Paul.³¹ Combining these quite disparate criteria yields an initial set of six authorship theories to be tested:

1. The original critical authorship theory of F.C. Baur and the older Tübingen School (four Paulines and twenty-one total authors).
2. A milder, less restrictive version of the old Tübingen School construct (four Paulines and nineteen total authors).
3. The current mainstream critical view (seven Paulines and eighteen total authors).³²
4. A modest refinement of the current mainstream view (seven Paulines and seventeen total authors).
5. A still largely mainstream authorship view that grants Colossians to Paul, per Kümmel (eight Paulines and fifteen total authors).
6. The traditional view of authorship (thirteen Paulines and nine total authors.)

These six authorship theories group the twenty-seven texts of the New Testament as follows:

³¹ This second insight is quite relevant because contemporary New Testament scholars vary quite widely in their views of the authorship of the New Testament.

³² Pace Mealand, "Positional Stylometry Reassessed," 267, who somewhat surprisingly replaced Philemon with 2 Thessalonians yielding Mealand's seven: Romans, I & 2 Corinthians, Galatians, Philippians, and 1 and 2 Thessalonians.

	Original Baur (21 Authors) ³³	Reduced Baur (19 Authors)	Modern Critical (18 Authors)	Modern Critical (17 Authors)	Modern Critical (15 Authors)	Trad. (9 Authors)
Matt.	Matt. Writer	Matt. Writer	Matt. Writer	Matt. Writer	Matt. Writer	Matthew
Mark	Mark	Mark	Mark	Mark	Mark	Mark
Luke	Lukan Writer	Lukan Writer	Lukan Writer	Lukan Writer	Lukan Writer	Luke
John	2nd Cent. Jn.	2nd Cent. Jn.	John, Redactors ³⁴	John, Redactors	John, Redactors	John
Acts	Lukan Writer	Lukan Writer	Lukan Writer	Lukan Writer	Lukan Writer	Luke
Rom.	Paul	Paul	Paul	Paul	Paul	Paul
1 Cor.	Paul	Paul	Paul	Paul	Paul	Paul
2 Cor	Paul	Paul	Paul	Paul	Paul	Paul
Gal.	Paul	Paul	Paul	Paul	Paul	Paul
Eph.	Eph. Writer ³⁵	Eph. Writer	Eph. Writer	Eph. Writer	Eph. Writer	Paul
Phil.	Phil Writer ³⁶	Paul	Paul	Paul	Paul	Paul
Col.	Col. Writer ³⁷	Col. Writer	Col. Writer	Col. Writer	Paul ³⁸	Paul
1 Thess.	1 Th. Writer ³⁹	1 Th. Writer	Paul	Paul	Paul	Paul
2 Th.	2 Th. Writer ⁴⁰	2 Th. Writer	2 Thess. Writer	2 Thess. Writer	2 Thess. Writer	Paul
1 Tim.	1 Tim. Writer ⁴¹	Pastoralist Paul ⁴²	1 Tim. Writer	Pastoralist Paul	Pastoralist Paul ⁴³	Paul
2 Tim.	2 Tim. Writer	Testamental Paul	2 Tim. Writer	Testament Paul	Pastoralist Paul	Paul
Titus	Titus Writer	Pastoralist Paul	Titus Writer	Pastoralist Paul	Pastoralist Paul	Paul
Philmn.	Phlmn. Writer ⁴⁴	Phlmn. Writer	Paul	Paul	Paul	Paul
Heb.	Heb. Writer	Heb. Writer	Heb. Writer	Heb. Writer	Heb. Writer	Unknown
James	James Writer	James Writer	James Writer	James Writer	James Writer	James
1 Peter	1 Peter Writer	1 Peter Writer	1 Peter Writer	1 Peter Writer	1 Peter Writer	Peter
2 Peter	2 Peter Writer	2 Peter Writer	2 Peter Writer	2 Peter Writer	2 Peter Writer	Peter
1 John	Jn. School	Jn. School	John, Redactors?	John, Redactors?	John, Others? ⁴⁵	John
2 John	Jn. School	Jn. School	John the Elder	John the Elder	John the Elder	John
3 John	Jn. School	Jn. School	John the Elder	John the Elder	John the Elder	John
Jude	Jude Writer	Jude Writer	Jude Writer	Jude Writer	Jude Writer	Jude
Rev.	1 st Cent. Jn.	1 st Cent. Jn.	Asia Minor Jn.	Asia Minor Jn.	Asia Minor Jn.	John

Table 5.3. Authorship Theories by Texts of the New Testament

³³ For Baur's evaluation of the nine Paulines he rejected, see Baur, *Paulus, der Apostel*, 417–99.

³⁴ Some scholars see core source material in John (that has been subsequently redacted) connecting a certain eyewitness John to Jesus (Howard, *The Gospel According to John*, 460).

³⁵ Baur (1845) followed Evanson (1792) and De Wette (1843) in weighing against the authenticity of Ephesians due to its perceived dependence on Colossians. Baur, *Paulus, der Apostel*, 417–18; Evanson, *The Dissonance of the Four Generally Received Evangelists*, 312–13; de Wette, *Kurze Erklärung der Briefe* 79.

³⁶ Though De Wette affirmed Philippians as Pauline beyond any dispute, Baur saw in it purported gnostic ideas. See especially Baur, *Paulus, der Apostel*, 458–64; de Wette, *Lehrbuch der historisch-kritischen Einleitung*, 267–68.

³⁷ Kümmel notes that the modern era of critical arguments against Pauline authenticity of Colossians began with Dibelius. Kümmel, *Introduction to the New Testament*, 140.

³⁸ Kümmel supports authenticity based on its style (Kümmel, *Introduction to the New Testament*, 341–2).

³⁹ Baur rejected 1 Thessalonians because it lacked originality. Baur, *Paulus, der Apostel*, 480–81.

⁴⁰ J.E.C. Schmidt (with K.C.L Schmidt) was the first to contest 2 Thessalonians due to perceived differences between the *parousia* of 1 and 2 Thessalonians (Schmidt, "Vermutungen über die beiden Briefe an die Thessalonicher," 380–86).

⁴¹ Schmidt doubted both epistles on the basis of aberrant vocabulary, and Schleiermacher doubted them because he viewed them as unknown to Polycarp as well as to the apostolic fathers. See Schmidt, *Historisch-Kritische Einleitung*, 257–67; Schleiermacher, *den sogenannten ersten Brief des Paulos an den Timotheos*, 16–19.

⁴² On the similarity of 1 Timothy and Titus, see Aageson, *Paul, the Pastoral Epistles, and the Early Church*, 87.

⁴³ Kümmel viewed 1 Timothy as the work of a single pseudepigraphist (Kümmel, *Introduction to the New Testament*, 385).

⁴⁴ Von Weizsäcker is one of the few that rejected this epistle (von Weizsäcker, *Das apostilische Zeitalter*, 183–84).

⁴⁵ Kümmel, however, views 1, 2, and 3 John to be by one hand (Kümmel, *Introduction to the New Testament*, 450).

5.4 Developing Testable Genre Theories

5.4.1 Stating the Genre Problem:

The third primary goal of this chapter requires that we develop separate, principled, *testable* theories of genre. To do so requires answering three prior questions regarding genre. First, *what is genre?* Second, how many distinct genres (or subgenres) exist in the NT? Third, as in the prior section regarding authorship, how many distinct genre sets must be built to “cover the waterfront”—to ensure that we capture the relevant set of alternatives? The first of these questions is doubtless the most critical, because the other two questions flow from it. To answer these questions requires that we consider the history of NT genre criticism, as well as identify problems endemic to its current state. The key to understanding the past history of NT genre criticism, in my view, lies in recognizing that more than anything else, it is a somewhat uneven history of the interaction between literary criticism, rhetorical criticism, and genre criticism. Understanding this interaction is a worthy exercise in its own right, and subsumes an expansive literature as evidenced by the reviews of Aune, Guelich, Blomberg, and Diehl.⁴⁶ Since the goal here is far narrower—to highlight only information relevant to developing *testable* theories of genre—I have quite selectively abstracted from that literature in the next section in order to build out a preliminary approach to genre. For the balance of this section, however, the focus will be upon the problems and limitations of the *current state* of NT genre criticism. Four such problems exist in my view: *consensus, depth, method and verification.*

⁴⁶ See Aune, *The New Testament in its Literary Environment*, 19–45; the survey by Guelich, “The Gospel Genre,” in Stuhlmacher, ed., *The Gospel and the Gospels*, 173–208; two surveys by Blomberg and his summary on genre from his major work on the Gospels, Blomberg, *The Historical Reliability of the Gospels*, 298–303; Blomberg, “New Testament Genre Criticism,” 40–49; Blomberg, “The Diversity of Literary Genres in the New Testament,” 505–32. Also see Diehl, “What is a ‘Gospel’?,” 171–99.

1. *Consensus*. Despite its long provenance, consensus still remains a problem in NT genre criticism. This must not be construed to mean that the last forty years of NT genre criticism has failed to yield a larger consensus, for surely it has. Indeed, two substantial syntheses have been achieved during that time. First, compared to forty years ago a larger percentage of NT scholars today would grant that the GNT displays at least the following genres: gospel, epistle, history, and apocalyptic—Aune’s four literary critical categories.⁴⁷ Second, the idea that the gospel genre is some sort of *sui generis*, with large elements of embedded Greco-Roman biography, has gained ground, though it may still lack consensus support. Despite these notable achievements, however, the current consensus regarding genre in the NT remains an uneven one. Agreement runs quite high, for instance, on epistolary membership, with less agreement on the classification of the gospel genre,⁴⁸ and even less regarding Acts as history. More profitable work, clearly, remains to be done.
2. *Depth*. The prior discussion of the uneven consensus of NT genre quite naturally raises another issue. While virtually no disagreement exists on which texts belong in which categories (i.e. *category separation*), can the same be said of *category depth*? That is, do Aune’s four categories adequately explain the diversity of genres and subgenres found in the NT as a whole? Specifically, while most would agree that classifying Revelation as apocalyptic is quite satisfying, does classifying fully twenty-one of the twenty-seven texts of

⁴⁷ See also Aune’s seminal study; Aune, *The New Testament in its Literary Environment*.

⁴⁸ The Gospel form may have proved difficult to nail down for another reason. For most of the twentieth century, the major methodological movements in New Testament scholarship assumed an extended period of literary composition, whether by compiling of sources, editing of independent units of oral tradition, or redaction of the subsequent product. By doing so, the intellectual frontispiece of twentieth-century New Testament scholarship shifted the locus of composition from a putative single author to the theological community. This shift left literary criticism, in particular, less certain of the gospel form, as opposed to a theory which conventionally (though not necessarily) assigned texts to a single author. An alternative explanation is just as theoretically possible: the Christ event was so rooted in history (biography) and yet so mythical in character that no existing hero legend form (Talbert) or historical biography form could contain it.

the NT under the same omnibus rubric of “epistle” leave us equally satisfied? Are we satisfied in classifying Philemon together with Romans, for instance—or Hebrews with 2 Timothy? Here we seek the fulfillment of the Ricoeurian ideal, the discovery of those literary genres that are truly irreducible. A next potential step in that direction is to return again to the field where Deissmann labored almost nine decades ago—the Greco-Roman epistle and letter-forms.

3. *Method.* Few would dispute that NT genre criticism, historically understood, is the dialectical interaction between literary criticism, rhetorical criticism, and genre criticism and their respective methodologies. Yet, a large number of methodological questions still swirl around this integrative dialectic. First, simply adopting the unmodified preexisting categories from either rhetorical or genre criticism and applying them to the NT is problematic given that (a) no real cross-disciplinary consensus exists on these categories, at least as applied to the New Testament, (b) many of these categories seem better suited to defining subsections of the texts (“books”) in the New Testament than the texts as a whole, (c) Greco-Roman oral rhetorical categories, for all their perceived advantages, fail to fully capture the extent to which the Hebreo-Palestinian culture among the other cultures of the ancient Middle East was a remarkably literary one,⁴⁹ and (d) quite a number of NT texts seem to defy any oral (or

⁴⁹ The flood of third-century rabbinic literature signaled by the Mishnah and Tosefta is not to be understood as an outpouring of literary activity due only to the exigencies of the diaspora. Rather, even in the apparent Jewish literary quiescence between the close of the *Tanakh* and the Mishnah, the Jewish sages (Chazai) reinforced the literary ethos of a very literary people by interpreting, expounding, and expanding on the *debar* of YHWH. By this activity, throughout their respective eras (the Zugot, Tannaim, Amoraim, Savoraim) the Chazai alongside the scribes (*sopherim*) formed a bridge between the *Tanakh* and the better known rabbinic literature. While their literary efforts quite naturally waxed and waned, what is quite clear is that among the cultures of the ancient near east, the Hebrews codified in their culture and cultus a place of authority for literature. Accordingly, in our view, rather than merely interrogating the Greco-Roman world for suitable NT subgenres, an exploration of Hebreo-Palestinian subgenres (nascent mishnaic *Shisha Sedarim*?) should be considered—especially with respect to the gospel form and the NT texts of Hebrews and James.

for that matter, literary) categorization at all.⁵⁰ However, there exists an even greater methodological problem. We find ourselves in an interregnum of sorts in NT studies. The well-understood long-term trajectory of New Testament scholarship as a whole—author to recipient to reader—has failed to integrate with a substantial contrapuntal trajectory, the shift from the dominance of diachronic methods to more synchronic, text-primary methods.⁵¹ Both trajectories meanwhile have cranked out their own methods, largely isolated from the other. The result, no working NT scholar would fail to confirm, is a fragmented methodological landscape. This raises two questions. First, have any other disciplines solved their own fragmentation problem? If so, we should take some notes! Second, might useful methodologies exist elsewhere to help clarify the outstanding issues we face, adding *categorical depth* in general and developing subgenres specifically?

4. *Verification*. In such a fragmented landscape some external method of verification is needed, one that can reign in the excesses of untethered theory on the one hand, while avoiding the gallingly positivistic, privileged and fundamentally naïve constructs that have often attached themselves to quantitative verification on the other.⁵² The nature of such a verificational

⁵⁰ Much of the work in rhetorical and genre criticism applied to the GNT has been properly criticized because its categories deal almost exclusively with *oral* genres. Despite Achtemeier's protest that the orality of the late antiquity was so pervasive that it knew nothing of the "silent, solitary reader" (Achtemeier, "*Omne verbum sonat*," 17), this was almost certainly *not* the case. Whatever else it may be, the NT is a *written* corpus, and truly endogenous categories for it must not solely presume it to be "speech written down."

⁵¹ The reason for this disjunction is not difficult to discover—the remarkable disinterest of postmodern interpretation in the text *as text*. This Roland Barthes telegraphed in his famous dictum, "The birth of the reader must be at the cost of the death of the Author" (Barthes, *Image, Music, Text*, 148). Against this neglect, see Stephen Fowl's essay, "Authorial Intention in the Interpretation of Scripture" in Green and Turner, eds., *Between Two Horizons*, 71–87. See also Vanhoozer, *Is There a Meaning in This Text?*, 201–80.

⁵² Agreeing with Michael Polanyi's optimism that science can be done without reductionism but not without commitments (Polanyi, *Personal Knowledge*, 59–61), in my view what is critically needed in addressing genre is some hyperordinating theory that makes sense of the particulars. Yet, it must also weave together the baseline contributions of its subdisciplines while reigning in their more fanciful notions. Contrary to the objections of some, Polanyi's approach is in no way negated by Hilbert's failed synthesis in theoretical mathematics or Gödel's seminal discovery in algorithmics. These scholars discovered limits to mathematical knowing in a way very compatible to Polanyi's critique of positivism (Polanyi, *Science, Faith, and Society*, 76). Mathematical positivists must yield only their Procrustean pretensions, not their legitimate strengths. On the other hand, neither should the relative worth of

method is critical because past attempts at external verification often have been less than successful.⁵³

To address the four critical questions regarding the current state of genre discussed above, I propose that all four can be substantially retired by focusing on the last two listed problems. That is, moving forward requires developing (a) a *derivational method* to enhance genre category depth and (b) a *verificational method* to test it. This is best done through two separate approaches, followed by a comparison of their results. In the first approach, I propose to integrate only the most secure contributions from literary theory—rhetorical criticism and genre criticism—to initially address the problem of *categorical depth*. This will yield an initial classification of all twenty-seven texts of the GNT in a way that is sensitive to the best contributions of each discipline. Once this is done the next step is to step back and qualitatively assess what kind of chimera has been built. In the second approach, the newly developed chimera will be domesticated, so to speak, by an independent endogenous theory of genre using synchronic co-textual insights informed by an external empirico-statistical method of verification. While that last sentence is admittedly densely worded, it will be unpacked in the next section.

an academic proposal be arbitrated merely by the consensus of like-minded dialog partners, but in some way external to their caprice. IBE holds the feet of both these groups to the fire.

⁵³ Here we paraphrase Pearson and Porter's apt insight that we can no longer subscribe to the fiction that an ancient archetypical theory of genre existed to allow us to verify our assignments of genre (Pearson and Porter, "The Genres of the New Testament," 136). In lieu of not receiving succor from contemporaneous Greco-Roman literary categories, the question then becomes, how can we verify our view of genre? This is the task of the next section.

5.4.2 Approach One: Integrating Literary, Rhetorical and Genre Theory

5.4.2.1 Literary Theory and Its Contribution to New Testament Genre

Contribution. Most early work in genre related to the NT, especially prior to the development of rhetorical and genre criticism, viewed genre in the GNT wholly in terms of literary theory or, at best, it viewed genre as a historically-situated literary category. The reason for this is patent. After the ground-clearing work by Deissmann in classifying NT texts as epistles⁵⁴ a work not intended to speak holistically to literary genre in the NT,⁵⁵ other scholars mistook his discovery as a method, and the die was cast to link NT texts to first- and second-century Greco-Roman literary genres. The problem with this lay in the brute fact that many of the texts in the NT passed through stages of orality—and perhaps extensive ones—before their literary incarnations. Despite this, what *literary theory* meant by genre became the *de facto* framework for understanding the genres of the NT in the early twentieth century. This dominance is traceable to three factors. First, nothing prevented a literary construal of genre from running the field in the early twentieth century given that rhetorico-critical, genre-critical, linguistic and especially sociolinguistic contributions to genre simply did not yet exist. Second, when the long dialectic regarding the genre of the Gospels (represented only federally by the views of Talbert, Aune, and Burridge) resulted in a larger consensus that the Gospels were *sui*

⁵⁴ Some trace the one-to-one correspondence between a text (book) of the New Testament and a given literary genre to Deissmann, but this misunderstands Deissmann's achievement. By comparing the NT texts to the newly discovered troves of papyri, Deissmann discovered that the NT epistles were not specialized literary constructs but quite typical examples of contemporaneous Greco-Roman letters. In so doing, Deissmann solidly situated the NT texts within the Greco-Roman world (Deissmann, *Light from the Ancient East*, 69–145). As a result of Deissmann's work it no longer seemed far-fetched that the putative authors of the NT quite naturally appropriated existing communication forms (epistles and letters) to achieve their semantic and persuasive ends.

⁵⁵ Deissmann's point was quite the opposite. What we now classify as the documentary papyri he termed *popular* texts. These were not literary in any real sense, because he viewed the central characteristic of literary texts to be intentional artistry. To confuse the two was according to Deissmann, "utterly perverse." Deissmann, *Light from the Ancient East*, 148–9.

*generis*⁵⁶ (with, arguably, strong elements of biography),⁵⁷ there was an informal sense that the tallest peak in NT genre studies had been surmounted.⁵⁸ With the status of the Gospels now somewhat less of a question mark, Aune's modest proposal, even if it failed to adequately address *category depth*, did quite adequately address *category separation*. Third, Aune's seminal study in 1987, whatever else it achieved, influenced a generation of New Testament scholars to remain within the relative safety of a four-category literary critical view of New Testament genre. Accordingly, the preponderance of scholars who have bothered to enumerate genres in the New Testament at all have done so largely within a literary framework. Thus, when Theissen and Boring enumerate two genres;⁵⁹ Marshall, Barton, and Dormeyer three genres;⁶⁰ Conzelmann, Blomberg, McKnight, Harrington, Klauck, Fee, Metzger, Ehrman, and many others four

⁵⁶ Aune in his important article in volume II of the *Gospel Perspectives* series (Aune, "What is Gospel?," 9–60) attacked Talbert's view of the Gospels as didactic mythic biography (Talbert, *What is a Gospel?*, 107) as did Burridge who viewed the Gospels as clear examples of biography (Burridge, *What are the Gospels?*). Wills was only slightly more nuanced, seeing the genre as "aretological biographies" (Wills, *The Quest of the Historical Gospel*, 10). The end result of this dialectic was an increasing consensus that although the gospel form is perhaps closest to a Greco-Roman biography, no contemporary Greco-Roman or Hebraic-Palestinian literary glass slipper really fits for the Gospels. Stephens' summary is memorable: "Many scholars have recognized the limitations of trying to force New Testament discourse into the pure generic categories of Aristotle or Quintillian." Stephens, *Annihilation or Renewal?*, 15.

⁵⁷ Votaw was the first to isolate strong touch-points between the gospels and Greco-Roman popular biography such as Philostratus' *Life of Apollonius*. Votaw, "The Gospels and Contemporary Biographies," 59–73.

⁵⁸ My own informal literature review demonstrates that this issue was hotly contested in the period from the 1970s into the early 1990s, peaking around 1990. While Diehl notes that derivational, newer analogical and speech act criticism of the gospel genre still continues unabated (Diehl, "What is a 'Gospel'?", 180–95), the essential modest consensus of the Gospels as a biographical *sui generis* still stands.

⁵⁹ Theissen's two basic categories include gospel and letter. He characterizes them as "literary forms." Theissen views the book of Acts and Apocalypse as a sort of subgenre; these are texts which found their way into the NT via their connection to the gospel form and the letter form respectively. Theissen, *The New Testament: History, Literature, Religion*, 8–13. In his more recent work, Theissen sees the literary genres in terms of three sequential historical phases: charismatic, pseudepigraphic and functional. Accordingly, he has added Hebrews to Acts and the Apocalypse as independent forms which constitute the functional phase of Canon formation (Theissen, *The New Testament: A Literary History*, 179–204). Boring understands the two literary genres as gospels and letters (Boring, *An Introduction to the New Testament*, 7).

⁶⁰ Marshall views the NT in terms of three literary genres: gospel, letters and apocalypses (Marshall, *New Testament Theology*, 21). Barton's perspective enumerates three genres relevant to the New Testament: apocalyptic, the Gospels and letters; Barton, *The Biblical World*, 129–48, 62–86, 87–202. Dormeyer, similarly, see the three genres as letters, gospels/Acts, and apocalyptic; Dormeyer, *The New Testament Among the Writings of Antiquity*, 205–55.

genres;⁶¹ it is to *literary* genres they refer. Even the more current work to classify NT texts outside Aune's now familiar fourfold complex is still largely defined within the boundaries laid down by literary theory.⁶²

Conclusion. On balance, it seems best to view the contribution of literary criticism as a fundamental division, a primary level of organization. That is, scarcely any improvement can be made to the categorical separation provided by Aune's categories. Accordingly it seems best to propose that these categories serve as our primary categories within which further subdivisions can and should be made.

5.4.2.2 Rhetorical Criticism and Its Contribution to NT Genre⁶³

Contributions: as a self-described tool to complement form and literary criticism, per Kennedy,⁶⁴ rhetorical criticism does indeed bring us into helpful conversation with forms of orality contemporary to the New Testament. Moreover, rhetorical criticism is doubtless correct in

⁶¹ Conzelmann's four genres are gospel, letter, historical monograph, and apocalypse. Oddly though, Conzelmann and Lindemann view the gospel as a type already available to the NT writers! See especially Conzelmann and Lindemann, *Interpreting the New Testament*, 26. Blomberg sees four literary genres at work in the NT, but multiple literary forms (Blomberg and Markley, *A Handbook of New Testament Exegesis*, 103–6). McKnight comments that "Four literary types comprise the New Testament." McKnight, *Introducing New Testament Interpretation*, 7. Harrington's enumeration of four genres is found in the context of his comments on literary criticism (Harrington, *Interpreting the New Testament*, 13). Klauck refers to his four categorizations as literary genres (Klauck, *The Religious Context of Early Christianity*, 448). See also Ehrman, *Lost Scriptures*, 156; Fee, *New Testament Exegesis*, 3; and Metzger, *The Canon of the New Testament*, vii. Compare all to Aune's categories (Aune, *The New Testament in its Literary Environment*, 13).

⁶² The recent more refined genre classifications of the Pastorals by Fiore and Johnson, for instance, fall under this head. See Fiore, *The Function of Personal Example*, 6; Johnson, *The First and Second Letters to Timothy*, 96–97.

⁶³ A short mention should be made that rhetorical criticism must be further leveraged for the insights it brings to understanding subgenres and further critiqued regarding its lack of controls and the breadth of its claims. For a comprehensive study of rhetorical criticism in the New Testament, see Kennedy, *New Testament Interpretation Through Rhetorical Criticism*, 3–156. For surveys of the literature of rhetorical criticism in the NT see Black, "Rhetorical Criticism and the New Testament," 77–92; Lambrecht, "Rhetorical Criticism and the New Testament," 239–53; Mack, *Rhetoric and the New Testament*, 25–48; Watson, *Invention, Arrangement, and Style*, 8–28. For the benefits and limitations of rhetorical criticism related to the genre of the GNT see the essays of Porter, Classen and Reed in Porter and Olbricht, eds., *Rhetoric and the New Testament*. Some more recent studies have attempted to use rhetorical criticism to aid in delegitimizing current political or ideological structures (Hester and Hester, *Rhetorics in the New Millennium*, passim). These latter efforts, unfortunately, largely seem to be celebrations of invention apart from sufficient controls. For a non-traditionally-written deconstruction of rhetorical and socio-rhetoric deconstructions, though still absent adequate controls, see Amador, *Academic Constraints in Rhetorical Criticism*, passim.

⁶⁴ See Kennedy's measured perspective, Kennedy, *New Testament Interpretation Through Rhetorical Criticism*, 3.

its observation that persuasion, the stock and trade of Greco-Roman rhetorical praxis, lies at the semantic core of many sections of texts within the NT. It seems clear, for instance, that the categories of judicial/forensic rhetoric are present to some extent in Galatians and Romans in particular,⁶⁵ and that sections of 2 Corinthians and doubtless Paul's multiple defenses in the latter chapters of Acts are signal examples of apologia.⁶⁶ Diatribe, moreover, clearly seems to have influenced the structure of Romans.⁶⁷ Other examples of Greco-Roman rhetoric within the texts of the NT seem either less prominent or comprise a substantially smaller percentage of the total information flow of the text.

Limitations. Despite the noteworthy examples above, with the exception of Romans and Galatians, none of the oral categories of Greco-Roman rhetoric seem to adequately capture the totality of the texts where they are found. More to the point, one is moved to wonder whether the categories of oral Greek rhetoric apply *at all* to either the gospel form⁶⁸ or especially to the more personal, occasional, and quite obviously non-oral texts that form much of the smaller epistles in the NT.⁶⁹ One questions whether, apart from perhaps Romans or Galatians or in a handful of

⁶⁵ Cf. the inaugural work of Hans Dieter Betz on Galatians (Betz, *Galatians: A Commentary*).

⁶⁶ The sequential defenses of Paul before the people (Acts 22), the Sanhedrin (Acts 23), Felix (Acts 24), Festus (Acts 25), and Agrippa (Acts 26) constitute one of the most sustained narrative examples of Greco-Roman apologia in the Koine.

⁶⁷ Song, echoing Stowers, notes that the general structure of Romans starting from 1:16 is consistent with the macro-structure of a diatribe (Song, *Reading Romans as a Diatribe*, 65). Song modestly expands Stower's list of diatribal sections to include 1:18-32, 2:1-5, 17-29; 3:1-18; 3:27-4:2; 4:3ff.; 6:1-23, 15-16; 7:7-25; 9:14-15, 19-21; 11:1-3, 11,13-25, and 14:4,10 See Stowers, *The Diatribe and Paul's Letter to the Romans*, 85-115; Song, *Reading Romans as a Diatribe*, 57. Porter largely concurs, arguing that Romans 5 also contains signature elements of diatribe (Porter, "The Argument of Romans 5," 665-77).

⁶⁸ As Kennedy observes, rhetorical critical discussions of the gospel genre are quite beside the point because only Luke shows any awareness of classical genres (Kennedy, *New Testament Interpretation Through Rhetorical Criticism*, 97).

⁶⁹ These texts fall far more easily within the boundaries of literary rather than oral/rhetorical theory. Even the protrepsis (πρότρεις) and paraenesis (παραινεσις) found in them, if indeed this distinction was understood at all in the first several centuries of the Common Era, have more of a documentary feel than a formal literary character.

other places in the NT, the authors/editors/redactors actually thought of structuring their writing in terms of Greco-Roman oratory at all.⁷⁰

Conclusion. The above limitations largely contraindicate using formal Greco-Roman rhetoric as a primary organizing category of NT genre. The unambiguous presence of formal epideictic, juridical/forensic, apologia, deliberative or other formal rhetorical elements within the NT texts may certainly be used as a *subclassification* of genre, to contribute *categorical depth* to our admittedly experimental classification. We must be cautious however, to govern the extent to which rhetorical categories contribute to *categorical depth* based on their prominence in the text (the total number of unambiguous instances) and the contribution of those elements to its overall information flow.

5.4.2.3 Genre Criticism and Its Contribution to NT Genre⁷¹

Contributions. Only two contributions of genre criticism need concern us here. First, Mikhail Bakhtin is to be credited with the landmark insight that primary (speech) genres give rise to secondary genres that absorb and digest primary genres – forming what could be termed hybrid genres.⁷² In focusing on genre within the NT, Blomberg concurs, stating more

⁷⁰ In this regard, Achtemeier states that oral performance can be understood even in the absence of any formal training on the part of either the writer or the auditor (Achtemeier, “*Omne verbum sonat*,” 9). This is clearly true, with the consequence that it is persuasive features of the language that persuade, echoed and likely intensified by well-understood rhetorical forms resident in the social repository of the language.

⁷¹ Frow boldly asserts that “genres actively generate and shape the knowledge of the world...[they] create effect of reality.” Frow, *Genre*, 2. This semantic function of genre echoes Halliday’s oft-repeated perspective that language construes meaning. It is for this reason we define genre as a shared social semiotic. NT scholars necessarily view genre more narrowly as the discourse and dialectic surrounding the classification of the texts of the NT into their most natural categories.

⁷² See Bakhtin, *Speech Genres and Other Late Essays*, 61–67. See also Jamieson and Campbell, “Rhetorical Hybrids,” 146–57. Bakhtin’s appropriation by biblical scholars, most notably in the series of SBL essays assembled by Boer, reflects the application of Bakhtin’s categories of monologic/dialogic narrative, chronotype, and polyphony to the biblical texts (Boer, *Bakhtin and Genre Theory*). Some of the interaction in Boer’s volume, such as the “didactic voice” in Mandolfo’s essay (Mandolfo, “Dialogic Form Criticism,” 73) begins to recapitulate language features that can be ascertained more quantitatively via the lexicogrammar or the semantics of the text or both. The almost anecdotal nature of many of these categories, however, seems to be far less useful than Bakhtin’s initial insight that genres have subgenres that are emergent from the context and, we would argue, from the cotext itself.

colloquially that genre categories in the NT are “mixes or combinations of genres.”⁷³ Second, genre criticism is increasingly open to interrogating socio-rhetorical contexts that give rise to genre.⁷⁴ With adequate controls in place, this trajectory can expand the locus of categorization from speaker to auditor as well as to occasion.⁷⁵ Such an expansion increases the dimensionality of the classification of genre and is to be welcomed.

Conclusion. Genre criticism gives us explicit permission to view genre flexibly as combinations of smaller forms that combine to create larger forms.⁷⁶ The implication of this level of flexibility is as intriguing as it is necessary. Just as simple units form more complex units so do primary categories give rise to novel *sui generum*. This permission (assuming appropriate controls) allows us to combine ancient literary rhetorical categories or subgenres together to (a) either become a primary organizing category or (b) constitute a subcategory as controlled by the same criteria as before, based on their prominence in the text (the total number of unambiguous instances), and the contribution of those elements to the overall information flow of the text.⁷⁷

With these three conclusions from literary criticism, rhetorical criticism and genre criticism, respectively, we can develop our guidelines for integrating them together to yield a more complete NT theory of genre:

1. Allow literary criticism to provide the primary (hyperordinating) categories.
2. Allow primary categories to combine together, if indicated, to form “mixes” of primary categories.

⁷³ Blomberg, “Genre in Recent New Testament Commentaries,” in Porter and Schnabel, eds., *On the Writing of New Testament Commentaries*, 75. Blomberg’s initial review of NT genre criticism is still a relevant mine for references related to NT genre criticism (Blomberg, “New Testament Genre Criticism,” 40–49).

⁷⁴ Robbins’ six categories of rhetorlect (precreation, wisdom, priestly, prophetic, miracle, and apocalyptic) deserve mention here. These are not genres *per se*, but rhetorical modes which were combined and recombined to form the major literary genres of the GNT (Robbins, “The Dialectical Nature of Early Christian Discourse,” 356).

⁷⁵ See Rhoads’ list of critical studies that have evolved from genre criticism (Rhoads, *Reading Mark*, 34–41).

⁷⁶ Wellek and Warren, citing Andre Jolles (Wellek and Warren, *Theory of Literature*, 246).

⁷⁷ Biber concurs with the notion of subgenres noting that “Some genres have several subclasses which are quite different from one another....” Biber, *Variation Across Speech and Writing*, 170. He then traces the ubiquity and interpretability of those subgenres throughout factors that he extracts using factor analysis (Biber, *Variation Across Speech and Writing*, 180–98).

3. Allow subordinating categories of genre to come from either oral (rhetorical) or written categories. Allow these to be mixes of categories as well.⁷⁸
4. Control the classification of texts into a single subcategory by *profile correspondence*, the display of a very similar set of features and/or subgenres.
5. Control the classification of genres so they do not become overly delicate (detailed). This insures that predominantly similar texts do not resolve into different categories.

These guidelines produce a number of self-similar hypothetical sets of genre categories.

One such construct is below:

	Closest Genre Category	NT Texts of That Genre
1	Gospel (<i>sui generis</i> . Closest to Greco-Roman biography with embedded elements such as pronouncement stories and miracles) ⁷⁹	Matthew, Mark, Luke
2	Gospel (<i>sui generis</i> . Greco-Roman biography with elements such as pronouncement stories and miracles; elements of φιλοσόφημα) ⁸⁰	John
3	History/Historiography (with embedded elements, i.e. miracles) ⁸¹	Acts
4	Epistle: Didactic/Forensic/Epideictic/Diatribic	Romans, Galatians
5	Epistle: Didactic/Epideictic	1 Corinthians
6	Epistle: Didactic/Epideictic/Apologia	2 Corinthians
7	Epistle: Didactic/Paraenesis ⁸²	Ephesians, ⁸³ Colossians, Philippians
8	Epistle: General/Elements of Apocalyptic	1-2 Thess., 1 Peter, Jude
9	Epistle: Hortatory/Prolepsis (Jewish Sermon?) ⁸⁴	Hebrews
10	Epistle: Hortatory/Ethical Paraenesis	James
11	Epistle: Personal Paraenesis with elements of <i>Mandata Principis</i> ⁸⁵	1 Timothy, Titus

⁷⁸ This is a restatement of Wellek and Warren's observation that complex literary forms (genre) develop out of simpler units (literary forms).

⁷⁹ These elements Burridge characterizes as a "family resemblance" (Burridge, *What are the Gospels?*, 243).

⁸⁰ Wills recognizes that John is different in some signal ways from the Synoptic Gospels. His extended discussion of hero elements in the Gospels (Wills, *The Quest of the Historical Gospel*, 23-50), however, is not fully convincing because it is unclear whether classifying the Fourth Gospel as a hero biography fits the available data better than a Greco-Roman biography and also because John's clear antithetical oppositions are in no way satisfied under this head. Wills' work, obliquely at least, has raised the issue of whether what we see in John's Gospel is *sui generis*. This may be the case given that John's antitheses fails to fit easily within Zoroastrian, Manichean, Greek dualist or any other ancient dichotomistic framework. Moreover, there is an almost peripatetic sense, with John functioning as the παιδεία to his readers—not perhaps in the στωά—but surely with his readers as οἱ μαθηταὶ αὐτοῦ. Hence, we propose a hybrid form for John: a Greco-Roman biography with elements of φιλοσόφημα—a philosophical treatise.

⁸¹ Or Roman historiography (Pearson and Porter, "The Genres of the New Testament," 146-48).

⁸² O'Brien notes that Ephesians is bifurcated: theological content followed by practical/paraenesis; O'Brien, *The Letter to the Ephesians*, 70-71. Yet we wonder if this doesn't apply to all seven critically accepted Paulines!

⁸³ Lincoln's comments are apt, "Ephesians resists clear-cut classification in terms of ancient epistolary and rhetorical categories." Lincoln, *Ephesians*, xxxvii.

⁸⁴ Swetnam sees it, correctly, as a sermon with some modest features of a letter (Swetnam, "On the Literary Genre of the 'Epistle' to the Hebrews," 261).

⁸⁵ Spicq was the first to conclude that the Pastorals were instructions from a superior based on parallels to the papyri (Spicq, *Saint Paul, les Épîtres Pastorales*, 1:35-36). Mitchell, however, views Spicq's argument as overdrawn given that Spicq relied on Rostovtzeff's translation of P.Tebt.703. Mitchell's own translation found it to be more of a simple memorandum (Mitchell, "P.Tebt 703 and the Genre of 1 Timothy," 254-57). Wolter, in contrast, views P.Tebt.703 as a *mandata principis*—although he also judged 1 Timothy and Titus as uncertain examples of the same (Wolter, *Die Pastoralbriefe*, 169). Regardless of how closely they follow a prescribed form, Johnson is doubtless correct that 1 Timothy and Titus are both personal and paraenetical. See especially Johnson, *The First and Second*

12	Epistle: <i>Mandata Principis</i> with elements of Personal Appeal	Philemon
13	Epistle: Testament with Elements of Paraenesis ⁸⁶	2 Timothy, 2 Peter
14	Epistle: Didactic with elements of a φιλοσόφημα	1-3 John
15	Apocalyptic ⁸⁷	Revelation

Table 5.4

Given Table 5.4, how shall we evaluate our literary/rhetorical/genre chimera? While there is much to appreciate in it, two aspects are quite unsatisfying. First, we seem to have produced, in fact, *too many categories*, perhaps in violation of guideline five. If we combine, say, the didactic epistles together, and the hortatory epistles together, we do reduce the total number of categories to eleven. But what controls are in place to tell us whether we have actually *improved* our categorization, rather than merely *simplified* it? There is a clear sense that there is too much subjectivity at play here. This in turn suggests that some clarifying construct may be missing. I will explore these missing features and their verification in the second approach below.

5.4.3 Approach Two: Proposing an Endogenous Theory of Genre

While integrating literary, rhetorical and genre criticism can produce genre categories (and many of them!), all three of these approaches are *extrinsic*; they develop their categories by some measure of similarity to other texts in the Greco-Roman world. In contrast, we propose a more balanced theory of genre, one that includes *extrinsic*, *paratrinisic*, and *intrinsic* determiners of genre. *Extrinsic* determiners, as we have seen, triangulate to external text traditions whether such traditions are literary or rhetorical in character. *Paratrinisic* determiners, in contrast, derive

Letters to Timothy, 96–97; Quinn, “Paraenesis and the Pastoral Epistles,” 193. In my view, these also display clear elements of a *mandata principis*. Cf. Aageson, *Paul, the Pastoral Epistles, and the Early Church*, 46.

⁸⁶ For justification on treating 2 Timothy separately from 1 Timothy and Titus see Murphy-O’Connor, “2 Timothy Contrasted,” 403–18 and Aageson, *Paul, the Pastoral Epistles, and the Early Church*, 87. For 2 Timothy as a testamentary letter see Kümmel, *Introduction to the New Testament*, 384; Bauckham, “Pseudo-Apostolic Letters,” 489; Quinn, “Luke and the Pastoral Epistles,” 194. On 2 Peter as a testamentary letter see Bauckham, “Pseudo-Apostolic Letters,” 493; Achtemeier et al., *Introducing the New Testament*, 528; Green, “Narrating the Gospel in 1 and 2 Peter,” 263. Whether these are a paraenesis with a testament or a paraenesis or a charge, see Smith, *Timothy’s Task, Paul’s Prospect*, passim. 2 Timothy and 2 Peter seem unique in the NT in both regards.

⁸⁷ Revelation by almost all accounts is a clear instance of the apocalyptic genre (Barton, *The Biblical World*, 132).

from the social context(s) that surround the instantiation of the text. By *intrinsic* determiners we mean that the categories must come upwards from the language of the NT itself. Such a categorization, if achievable, would be closer to an endogenous categorization for genre in the NT—one in which important determiners are far less likely to be absent (exogenous) from our view.⁸⁸ Second, and as before, we want to allow such a categorization to yield hybrid categories at even the most primary levels.⁸⁹ Third, as noted above, we seek a method external to our theory to test the validity of our categorization, including a method to discover the relative weighting of these three determiners.

Such a mandate is, indeed, theoretically elegant – but how can we achieve it? In a very noteworthy development, a handful of scholars have begun to lay the groundwork for this kind of work by casting genre in a different light. For these scholars, what we term literary genre is an *ex post facto* classification of sorts, an epiphenomenon that points to far more primordial forces at work—forces that are simultaneously historical, sociological and linguistic in character. For them, the genre of a text must be tied more closely to both the *context of culture and the context of situation* (we define this below) in which it arose, and that context has necessary implications for the linguistic choices made by the author/editor/redactor in constructing the text.

Accordingly, such scholars have attempted to construct genre categories using extra-literary and extra-rhetorical phenomena. Of particular interest are scholars who have suggested or inferred that what we monolithically view as genre may itself be composed of different linguistically and

⁸⁸ NT scholarship, influenced by the prior emergence of diachronic compared to synchronic analytics, has accordingly approached this topic in the reverse order to what I suggest here. What I propose parallels Hirsch's distinction between intrinsic and extrinsic genre. Hirsch argues that works are better interpreted by intrinsic material (the utterance itself) rather than extrinsic (the traditions and conventions the speaker relies upon). In Hirsch's words, "to know the intrinsic genre and the word sequence is to know almost everything." Hirsch, *Validity in Interpretation*, 88.

⁸⁹ Here Meeks astutely states that "The differences between the New Testament and the literary works of the Golden Age were often so great that...the Christian documents were put in a class by themselves" (Meeks, "Forward," 7-8).

contextually located subcomponents that constitutively give rise to what we recognize as literary genre today. These efforts can be seen nascently in the genre categories of Pearson and Porter,⁹⁰ quite explicitly in the work of J.R. Martin and D. Rose,⁹¹ and specifically in the time and addressee dimensions of the register profile developed for the NT quite recently by Pitts.⁹²

Independently from Pitts, but quite parallel to his thinking, I propose here that linguistics (linguistic measures across the rank scale) as well as semantic measures informed by *a posteriori* quantitative analytics, must play a crucial role alongside older historico-literary/rhetorico/genre critical and more recent sociolinguistic discussions of genre. The perspective I propose is predicated upon three central ideas. First, if genre is to be truly sensitive to the NT in its historic setting, it must also draw upon and link to cotextual *and* contextual phenomena from that era.⁹³ Second, and even more specifically, however one views genre, all culture have *shared social semiotics*, culturally understood constructs that the author/compiler/redactor can draw upon as demanded by the *context of situation*. Third, whereas historico-literary discussions of genre have hinged almost exclusively upon culture-level (e.g. Greco-Roman) phenomena while largely ignoring the contextual determiners (the context of situation), and whereas sociolinguistically informed discussions have sought to correct this oversight, insights regarding genre can also be induced via the abductive interplay between a principled *a priori* experimental design as

⁹⁰ See Pearson and Porter, "The Genres of the New Testament," 131–65.

⁹¹ Martin and Rose induce genres such as stories, histories, reports, explanations and procedures up from the text using tenor and appraisal theory in SFL and Martin's own discourse semantics. See especially Martin and Rose, *Genre Relations*, 21–37.

⁹² Pitts' work is influenced by Bell's audience design model for style-shift applied to the Pauline corpus based on Halliday's register metafunctions and Biber's situational components (Pitts, "Style and Pseudonymity in Pauline Scholarship," 122–30).

⁹³ That is, texts are not written in a vacuum. By the *cotextual* setting I mean that a text is written in conscious relationship to similar texts. The *contextual setting* is the social milieu that called forth the text. It includes the occasion, audience, and background of the theological community and the like. Hirsch stated a roughly equivalent notion nonlinguistically, his concept of *intrinsic genre*, the controlling idea of the text (Hirsch, *Validity in Interpretation*, 78–79).

informed by *a posteriori* quantitative analytics. To unpack this (second!) admittedly turgid description, we define each of these terms below.

Register is a formal linguistic construct. It reflects a restricted set of choices adopted by the sender to assist in the semiotic exchange—communicating so as to be best understood.⁹⁴ Said even more linguistically, the author/compiler/redactor (henceforth the sender) grammaticalizes (actualizes) her linguistic choices to adapt to the social (and in terms of the NT texts the theological) *situation*⁹⁵ of her readers or auditors. Second, to facilitate the semiotic exchange senders not only adapt their language choices (register) to their recipients, but also consciously employ culturally and sociologically *shared social semiotics*⁹⁶ (communication forms or vehicles) for doing so. These shared social semiotics are what we mean linguistically by genre. Contemporary shared social semiotics in the first several centuries of the Common Era include a complex of oral and written forms, many of which parallel quite commonly accepted literary or rhetorical categories or subcategories (i.e. paraenesis, diatribe, testament, family letter, *mandata principis*, and the like). These vehicles, in turn, further constrain the lexico-grammatical choices

⁹⁴ Disagreement exists about the exact distinction between register and genre. Halliday's definition of register (variation according to use) is helpfully expanded by Reed as the "linguistic expressions of doing different types of social activities by social-groups" (Reed, *A Discourse Analysis of Philippians*, 54). In contrast to Reed who sees no strong distinction between register and genre, our position affirms a clear distinction between the two. Register is an accommodation of the sender/speaker/writer to the receiver/auditor/reader which assists in the semiotic exchange. This takes the form of a pruned or more limited system of lexico-grammatical choices. Accordingly, we adopt one register when we order from a menu at an expensive restaurant and quite another when we cheer our favorite team at a ball game. In distinction to register, genre is the adoption of a preexisting social semiotic – a vehicle – for facilitating that exchange. Using a contemporary example, when we appropriate a certain meter in writing a love sonnet, or adopt the accepted letter form written to a high ranking official, we are using genre. Genre necessarily collapses certain ranges of choices within the system network—in some instances completely. Compare to Martin and Rose (Martin and Rose, *Genre Relations*, 6) who view genres as "recurrent configurations of meanings...that enact the social practices of a given culture."

⁹⁵ See Halliday's cline of instantiation, in which the *situation type* (which itself is situated between the context of situation and the context of culture) yields a repertoire of registers (Halliday and Matthiessen, *An Introduction to Functional Grammar*, 28).

⁹⁶ Halliday characterizes language itself as one form of a social semiotic; Halliday, *On Language and Linguistics*, 424. By adding the word *shared* here I seek to stress that communication vehicles (genres) in language are well-recognized within a culture.

of the sender. Third, the *context of situation*⁹⁷ is the entire immediate environment in which a text is formed or instantiated.⁹⁸ For brevity, I will sacrifice some precision and assume a rough congruence between the linguist's *context of situation* and the theologian's *Sitz im Leben* – including all the rich sociological frameworks such a congruence implies. Fourth, the combination of *a priori* analytics with *a posteriori* analytics not only implies that a principled *a priori* experimental design is in place to analyze language data, but also that such a design is *re-entrant*. That is, the results of the initial analyses are used to further inform the experimental design abductively. This in turn yields *a posteriori* enhancements to the original design, or in our case the testing of our genre theories. This historico/socio/linguistic definition of genre (or equivalently this cultural/contextual/cotextual definition) coupled with its analysis has three primary advantages. First, it is not anachronistic. It views genre in terms of different types of shared social semiotics indigenous to the first several centuries of the Christian era. Second, linguistically, humans grammaticalize their meanings quite differently under different contexts, such as when the auditor or readers are individuals rather than a group. Third, it hypothesizes that the *Sitz im Leben* of the hearers, and hence the senders, may be quite dramatically influenced by whether the auditors or readers come from a Hellenistic or Jewish culture or a mix of both. It should be noted that similar efforts to develop a formal sociolinguistic definition of genre (albeit not a quantitatively informed one) include the important recent foundational work of Biber,⁹⁹

⁹⁷ “Every act of meaning,” comments Halliday, “has a context of situation, an environment in which it is performed.” Halliday, *On Grammar*, 201.

⁹⁸ In SFL, language operates in contexts, with the context of a text within a context of situation, and the system of language itself a context of culture (Halliday and Matthiessen, *An Introduction to Functional Grammar*, 28).

⁹⁹ Biber argues that registers necessarily adapt to the social contexts (Biber and Finegan, *Sociolinguistic Perspectives on Register*).

Bell,¹⁰⁰ and Martin.¹⁰¹ Our categories are largely driven by the same concerns. Two final comments should be made relevant to uncovering truly endogenous genre categories—categories natural to the language of the NT, its social context(s) and its times.¹⁰² First, the *order* of the three characteristics above is not accidental; it recapitulates the phenomena of language itself. Genre is (a) instantiated in language by the author/editor/redactor, (b) addressed into a social context often via a shared social semiotic, (c) received by an auditor/reader and (d) made more salient by congruence with shared social semiotics in the larger culture. Second, this linguistic/sender/receiver view of genre is distinct from traditional views of genre. It is synchronic, rooted first in language because in the final sense our *ex post facto* literary categorizations (extrinsic genre) depend first upon the actualized choices of the author/editor/redactor of the text. In the fullest sense the choice of genre was first intrinsically the sender's before it ever became ours as the recipient. It is indeed an unfortunate artifact of history that the academic discourse of genre has proceeded in the opposite direction.

This theoretical background provides a framework for constructing a theory of genre and applying it to the New Testament. Four principles follow from this discussion. First, per our

¹⁰⁰ Bell recognizes and formalizes what has become a growing conception among linguists that extralinguistic (contextual) factors contribute markedly to language (cotextual) variation. Specifically, Bell demonstrates that speakers shift their style based on their audience. See especially Bell, "Language Style as Audience Design," 146. Such shifts are not limited to audience only but include, without limitation, topic shifts per Rickford and McNair-Knox, "Addressee- and Topic-Influenced Style Shift," 235–76 as well as race. While Bell's insights are welcomed, the fact that audience differences quite markedly affect lexicogrammatical choice has long been recognized. In his 1955 Tyndale lecture, Guthrie said the same quite memorably, "We should not expect Paul to write in the same vein to his intimate associates as to Christian communities." Guthrie, *The Pastoral Epistles and the Mind of Paul*, 27.

¹⁰¹ Martin's construct of genre and register (Martin, *English Text*, 496) adapts and enhances Halliday's diagram of stratification (Halliday and Matthiessen, *An Introduction to Functional Grammar*, 25).

¹⁰² The first two characteristics of our definition (natural to the language of the NT and its social context) maps to Hirsch's concept of intrinsic genre. It should be noted that our definition of extrinsic genre is quite different than Hirsch's. Our third characteristic ("its times") is *extrinsic* because it is extrinsic to the text. It is a diachronic characteristic which maps to the traditional diachronic categories of critical theory which itself includes literary, rhetorical and genre theory as subdisciplines. It does *not* map to Hirsch's extrinsic concept of genre which he regards simply as "a wrong guess"—a binary opposition to intrinsic genre (Hirsch, *Validity in Interpretation*, 88–89). See also the discussion by Pearson and Porter concerning the same (Pearson and Porter, "The Genres of the New Testament," 131–33).

guidelines, Aune's four categories will serve as our hyperordinating categories. This we term literary *type*. Second, by interrogating the Koine of the GNT, especially the epistles, it becomes readily apparent that an important binary choice in the system network of the Koine relates to *number*. The decision to promote number as our second organizing category has precedent given the longstanding observation made by Parry, Hitchcock, Simpson, and others that when Paul speaks to close associates, the *tenor* of his language changes.¹⁰³ Third, the *audience* addressed, whether it be Hebraic-Palestinian or Greco-Roman, doubtless shifts the lexicogrammatical as well as the semantic choices of the author/editor/redactor. This becomes our third and final organizing category.¹⁰⁴ While schemas other than our *type/number/audience* taxonomy are of course possible,¹⁰⁵ at least this schema better reflects the *intrinsic* and *paratrinic* features of the language of the New Testament than the *extrinsic* four-fold schema of literary theory alone. Fourth, there remains one complication. Even with our type/number/audience schema, it is still unclear in some instances how to categorize certain GNT texts. Is Hebrews, for instance, a Jewish sermon or an odd type of corporate didactic or hortatory epistle? If the former, it becomes its own category, if the latter it becomes possible to group it with Romans and other instances of the corporate didactic epistle form. To respect this uncertainty, it is necessary to test the texts in various possible genre "slots," and then allow the mathematics to abductively weigh in on which

¹⁰³ The terminology chosen here is not accidental. In SFL, shifts in the number of subjects addressed as well as the mood and manner (tenor) in which they are addressed fall within the interpersonal metafunction—the tenor of the language. See Halliday and Matthiessen, *An Introduction to Functional Grammar*, 106–67.

¹⁰⁴ Trobisch argues that $\Phi 46$ was organized by addressee, with the texts to theological communities preceding letters to individuals. If Trobisch is at all correct, not only is audience further confirmed as a natural organizing category for genre in general, it becomes uncontestedly situated as a natural thought category in the early Christian Era (Trobisch, *Paul's Letter Collection*, 52–54).

¹⁰⁵ Reed suggests that genre can be identified via five aspects of discourse: (1) subject matter (semantic content), (2) situation type (context of situation), (3) participant roles, (4) mode (persuasive, explanatory or imperatival) and (5) medium (spoken or written). The first four of these seem to be viable candidates for expanding or sharpening our view of genre. Reed, "Modern Linguistics in Historical Criticism" in Porter and Carson, eds., *Linguistics and the New Testament*, 40–41.

categorization performs best empirically. Following these principles, we have generated four distinct categorizations of genre.¹⁰⁶ The first is grouped by the richest combination of type, number, and audience elements. It yields thirteen groupings. We then successively simplified the type and audience dimensions to yield the remaining three genre categorizations as detailed in Table 5.5 below.¹⁰⁷

GNT Books by Genre (Following a Type / Audience / Number Typology)				
Closest Genre Category	Genre (13)¹⁰⁸	Genre (12)	Genre (10)	Genre (8)
Gospel: <i>sui generis</i> . Closest to Greco-Roman biography w/ traditional material (pronouncement stories, miracles).	Matt, Mark, Luke	Matt, Mark, Luke	Matt, Mark, Luke	Matt, Mark, Luke, John
Gospel: <i>sui generis</i> . Closest to Greco-Roman biography w/traditional material and elements of φιλοσόφημα)	John	John	John	N/A
History/Historiography (with embedded form critical elements such as miracles)	Acts	Acts	Acts	Acts
Epistle Corporate: Didactic/Diatribic	Romans, 1-2 Cor, Gal, Phil	Romans, 1-2 Cor, Gal, Phil	Romans, 1-2 Cor, Gal, Phil, Heb	Romans, 1-2 Cor, Gal, Phil, Heb
Epistle Corporate: Didactic/Paraenesis	Eph, Col	NA	NA	NA
Epistle Corporate: General	1-2 Thess, 1-2 Peter, ¹⁰⁹ Jude	Eph, Col, 1-2 Thess, 1-2 Peter, Jude	Eph, Col, 1-2 Thess, 1-2 Peter, Jude	Eph, Col, 1-2 Thess, 1-2 Peter, Jude
Epistle Corporate: Hortatory (Sermon?)	Hebrews	Hebrews	NA	N/A
Epistle Corporate: Jewish Paraenesis ¹¹⁰ (Ethical Paraenesis?)	James	James	James	James
Letter ¹¹¹ Individual: Paraenesis or <i>Mandata Principis</i>	1 Tim, Titus	1 Tim, Titus	1-2 Tim, Titus	1-2 Tim, Tit., Phlm

¹⁰⁶ These three dimensions, in fact, create a set of genre possibilities that are the product of their individual dimensions. For example, let's assume there are twenty types and subtypes of genre in the first century, by six audiences (the combinations of Jewish, Gentile, and Palestinian) by two numbers (corporate and individual). These three dimensions yield 120 possible type/audience/number combinations, most of which are not in evidence in the GNT.

¹⁰⁷ From now on I will conform my terminology to reflect pseudepigraphal convention and refer to genre *categories* as genre *theories*.

¹⁰⁸ Our synthetic approach integrating literary theory, genre theory and SFL has both differences from but even greater congruencies with the Pauline register profile developed independently by Pitts (Pitts, "Style and Pseudonymity in Pauline Scholarship," 133-38). Pitts' register configuration is more sociolinguistic in nature. Mine is more explicitly literary to reflect the diachronic trajectory of NT genre studies. Future work is planned to test the Pitts' five Pauline register categories empirically.

¹⁰⁹ Achtemeier's observation that 2 Peter has elements of testament is worthy of further investigation. In this case it would be grouped with 2 Timothy (Achtemeier et al., *Introducing the New Testament*, 528).

¹¹⁰ Given its address, "to the twelve tribes," and its tenor, a Jewish audience seems to be in view.

¹¹¹ Deissmann sees the essential difference between the epistle and letter to be public and for posterity vs. private and personal (Deissmann, *Light from the Ancient East*, 228-30). In his analysis all of the letters of Paul were not

Letter Individual: Elements of <i>Mandata Principis</i> /Personal Appeal	Philemon	Philemon	Philemon	<i>N/A</i>
Letter Individual: Testament with Elements of Paraenesis	2 Tim.	2 Tim.	<i>N/A</i>	<i>N/A</i>
Letter Individual: Appeal/Elements of φιλοσόφημα	1-3 John	1-3 John	1-3 John	1-3 John
Apocalyptic: Jewish	Revelation	Revelation	Revelation	Revelation

Table 5.5

In sum, now that Section 5.3 has yielded six testable authorship theories and Section 5.4 has yielded four testable authorship theories, the third part of this study (praxis) can now commence. Two final, related objections should be addressed before beginning the analysis. Each will be addressed in turn.

First, how *adequate* are the four initial genre theories at getting to the heart of the notion of genre? (We assume here that the authorship theories, drawn as they are directly from the history of NT scholarship, are not under serious question.)¹¹² Said alternatively, have these genre theories “hit the mark” or are these notions of genre “mixed-up” with other components of sociolect? It is quite easy to give a direct—though perhaps unsatisfying—answer to that question. Given that the next four chapters are the very first cycle through the abductive design, the research *cannot yet abductively weigh in on the goodness or purity of these genre theories*. The important thing, per the discussion of abduction in Section 2.2, is that we define a starting stance, justify a theory as precisely as possible, and *start*—so that abduction can iteratively weigh in on the explanatory power of these starting definitions.

Second, given that no data can yet be brought forward to verify that the initial genre categories are “pure” versions of genre, what justifies our calling these four genre theories *genre*

intended for posterity, and hence were true letters (Deissmann, *Light from the Ancient East*, 240). Because discerning authorial intention is problematic, I have simplified the definition to *number*: audience plurality vs. singularity.

¹¹² The assumption here is that the six authorship theories are drawn from the history of NT scholarship and are not in serious question.

rather than *sociolect*? The answer is that, as defined, there has been a principled effort to either incorporate previously identified sociolectic components into our notion of genre (such as audience in terms of culture and audience in terms of number) or hold other confounding sociolectic subcomponents (e.g. date of composition) constant. Incorporating previously hypothesized sources of sociolectic variance (which we assert to be part of genre) inside of genre and holding external sources of sociolectic variance constant justifies the claim that these four theories can be termed, at least initially, to be theories of genre rather than some racemic mix of sociolectic components. Given these considerations the data in Chapters Six-Nine will use the terms genre (not sociolect) and authorship (not idiolect) consistent with working definitions of stylistic variation presented in Section 2.4.3.

6 Results Part I: Testing GNT Authorship and Genre Using Univariate Analysis

In this, the first chapter of empirical results, two tasks will be undertaken. First, I will discuss how the GNT was prepared for analysis. Second, the GNT will be analyzed in terms of the authorship and genre theories just described using both univariate measures of nonparametric analysis (NPA) and univariate measures from information theory (IT).

6.1 Preparing the GNT for Analysis

Performing statistical linguistic analyses of the GNT requires both a well-designed database architecture and an appropriate suite of analytical software. To achieve these ends, I founded in 2007 the Integrative Greek New Testament Project (IGNTP), a privately-funded research initiative.¹ The *mission* of the IGNTP is to meaningfully contribute to longstanding issues in NT scholarship (e.g. pseudepigraphy, authorship, the Synoptic Problem, and the like) using statistical linguistics. The *focus* of the IGNTP involves implementing four intersecting research activities: (1) creating principled, theoretically-informed linguistic and extralinguistic² probes into the structures, systems, strata and functions³ of the Koine in general and the GNT specifically,⁴ (2) analyzing those texts descriptively, inferentially, structurally, systemically, decompositionally and causally using both *a priori* and *a posteriori* methods, 3) interrogating the results using multiple validation methods and abductively modifying the working experimental

¹ I wish to gratefully acknowledge the technical and programming support provided to the IGNTP by Decision Support Sciences, a marketing science consulting firm specializing in business analytics.

² Our initial focus has been upon building *lexicogrammatical* probes. The term, coined by Halliday, reflects Halliday's apt notion that lexis and grammar exist as a continuum (Halliday and Matthiessen, *An Introduction to Functional Grammar*, 43–45).

³ *Structure* refers to the syntagmatic cline of linguistic rank, while *system* refers to the multiple paradigmatic systems (choice networks) within any given language, in our case the Koine of the GNT.

⁴ The probes range not only across the levels of linguistic rank (e.g. words, word groups, clauses etc.) but each probe exists in multiple forms; they are sorted in frequency as well as by multiple measures of variation. Variation probes include language forms categorized by adjusted standardized residual, and various other forms of standardized deviates per Haberman, "The Analysis of Residuals," 206–13.

design as necessary, and (4) publishing the results. The *primary methodology* of the IGNTP can be summarized in five propositions. First, it is given that the GNT, in common with all language corpora, presents to the reader/auditor complex patterns of language variation. Second, the sources or components of the total variation have been well discussed (per the working definitions of stylistic variation in Section 2.4.3) and constitute a relatively small number of hypotheses to be tested, such as variation due to date, specific historical setting, audience, occasion (*Sitz im Leben*), and the like. Third, these putative sources of variation separately or in combination can be added as hypotheses to the database via various methods of encoding. Fourth, through the use of multiple and mutually reinforcing multivariate models the variance structure of the GNT can be resolved into its constituent components (main effects, interactions, latent variables, etc.) and the contribution of those components to the overall structure and systems of the GNT can be quantified. Fifth, the magnitude of the now disentangled components, their covariance relations, and the aptness of their fit can be used to objectively arbitrate between the various theories. The main product of the project is the Integrative Greek New Testament (IGNT) itself, a multiple table statistical database⁵ of the Greek NT where each row (record) of the database is one of the 138,019 words in the text of the GNT, and each column (field) is one of the probes discussed above.⁶ Currently, the IGNT database has achieved the following

⁵ The IGNT is not a TEI-compliant database, because TEI databases are not natively readable by commercially available statistical software. Accordingly, the IGNT has been implemented as an SPSS SAV file. This decision allowed us to create and test linguistic probes far more quickly than would have been possible with a TEI-compliant databases.

⁶ Within the main repository of IGNT the probes have been organized into two general categories, extralinguistic probes and linguistic probes. The extralinguistic categories currently include traditional grammar, lexical priming, collocations/colligations and a variety of developmental fields. The linguistic categories are derived from systemic functional linguistic theory and are probes that vary across the scale of rank in the GNT (i.e. word, word group, clause, and clause complex) as well as include semantics and discourse fields. Structurally, the IGNT is actually a family of databases: a main repository database and analytical sub-databases of those fields organized by analytical criteria. The sub-databases exist only for pragmatic reasons, since analytics run more rapidly on the smaller databases.

milestones: (a) full integration of five annotated eclectic texts of Greek NT into the IGNT, including GramCord™, AGNT, Logos, MorphGNT, and OpenText,⁷ (b) creation and verification of 4,086 fields⁸ including approximately 2,800 linguistic probes derived largely from systemic functional theory (the rank of the clause and below), (c) creation of ~800 probes from traditional grammar to serve as controls, (d) creation and verification of ~200 probes above the rank of the clause, (e) creation and verification of ~200 probes from language strata other than lexicogrammar including semantic domains, and (f) acquisition and assembly of an analytical software suite to analyze the IGNT. This suite currently includes SPSS Statistics (formerly IBM-PASW), the R programming language, PositionSolve™ and PrefSolve™ (the latter two are Decision Support Sciences' multivariate visualization and simulation modeling software respectively⁹) and various custom software tools developed by the IGNTP.

⁷ Special thanks are due to Paul Miller, Executive Director of the GramCord Institute for providing the GramCord GNT, Barbara and Timothy Friberg for their inaugural work on the Analytical Greek New Testament (a project now overseen by John Hughes of the AGNT project who provided the IGNTP with the AGNT), Logos Bible Software for access to the Logos GNT, Drs. Stanley E. Porter and Matthew Brook O'Donnell for access to the OpenText GNT, and James Tauber and Ulrik Sandborg-Petersen for creating online access to the MorphGNT.

⁸ The current working revision of the IGNT is Rev 85.

⁹ Because of severe table size limitations in SPSS (only 1000 rows or columns are allowed in a cross tabulation) we enhanced PrefSolve™'s cross tabulation capabilities to handle large tables, perform bootstrapping, and calculate a larger suite of information theory-based statistics.

6.2 Interpreting Contingency Tables and Their Statistics: A Primer

To appropriately interpret contingency tables it is necessary to recognize that contingency table statistics differ in their robustness to comparisons across tables that vary markedly in size or sparseness (the average expected counts per cell). This is especially relevant in language analysis because language measures tend to vary greatly in the natural number of categories they present for analysis. Major semantic domains in Greek, for instance, present only 93 categories for analysis, whereas inflected lexemes present 17,736. The experimental design for this study, therefore, generated contingency tables where the total number of cells per table varied over two magnitudes of order.¹⁰ To mitigate both well-recognized and less well-recognized sources of statistical and interpretive error six steps were implemented.

1. Normed measures of association were selected that explicitly adjust to table size: Cramér's V, Goodman and Kruskal's tau and the Uncertainty Coefficient. Henceforth, the latter two statistics will be abbreviated as G&K's tau and UC respectively.
2. Tables were adjusted to be closer to the same size by *binning* or *partitioning*. Binning is the simple practice of recombining into a final category language measures (e.g. specific lemmas) that occur less than a cutoff number of times. In this work frequency cutoffs of 5, 10, 25, 50, 100 and 200 times were employed for this purpose. Partitioning is the practice of separating a subset of the forms of that specific language measure (e.g. lemmas) from the remainder of the forms of that language measure. In this study we partitioned each language measure *by frequency* (e.g. sorting by decreasing frequency and including only the forms that contributed to the top N% of *instances* by frequency

¹⁰ If left unmodified, the table sizes drawn from the GNT for these measures vary from 744 cells (93 rows of major semantic domains by 8 columns of genre) to 478,872 cells (17,736 rows of inflected lexemes by 27 books of the GNT).

cumulatively) and *by category* (e.g. sorting by decreasing frequency and dividing each partition by the top N% of categories). From now on these two types of partition strategies will be referred to as *instance* or *category partitioning* respectively. Whether we partitioned by instance or category, we split the data into deciles (ten partitions) or quartiles (four partitions). This either yielded identical numbers of frequencies per table (in the case of instance partitions) or tables that varied by only a half a magnitude of order in number of cells (in the case of bins) while improving expected counts per cell.¹¹

3. The effect of bin size and partitioning on the *profile* of association measures were explored. A *profile* refers to the relative shape of the data when it is graphed. By comparing profiles of different bin and partition sizes we explored whether low-frequency forms differentially contributed to authorship or genre.¹²
4. Unbiased estimates of Cramér's V, G&K's tau, and UC were developed via bootstrapping. These yielded more robust estimates for many statistics, including those used here.¹³
5. Confidence intervals (CI's)¹⁴ for Cramér's V, G&K's tau and UC were calculated via bootstrapping.¹⁵

¹¹ The sole exception is major semantic domains which have only 93 actual disambiguated categories and necessarily produce a much smaller table.

¹² Doing so fully addresses the reasonable objection that binning introduces an effect that differentially attenuates the magnitude of the association due to authorship. The data demonstrated, in fact, that the overall profiles were remarkably stable. See, for example, the "UC by bins" worksheet at the following permalink:
ftp://decisionsupportsciences/stat_ling/Association_By_Book_4Genre_6Authorship_Functional_Bootstrap.xlsx

¹³ It does so by estimating the shape of the sampling distribution of the statistic itself rather than assuming that the data fits an idealized distribution as in traditional Fisherian statistics. An excellent monograph on bootstrapping can be inspected in the QASS Series, Mooney and Duval, *Bootstrapping: A Nonparametric Approach*.

¹⁴ Calculating CI's via bootstrapping is much preferred to calculating CI's via ASE (adjusted standardized error). Jeong et al. forthrightly state that "When the sample size is small and the contingency table size is large, the bootstrap method is more powerful than the other methods." Jeong et al., "Bootstrap Tests for Independence," 628.

¹⁵ Bootstraps were executed to a 95% CI using 100 samples via the percentile method and nonstratified sampling. Bootstrapping is superior to standard (ASE) methods of calculating CI's. See Mooney and Duval, *Bootstrapping: A Nonparametric Approach*, 4-9.

6. Monte Carlo Simulations¹⁶ (MCS) for Cramér's V were executed. The rationale for this was to address the slight bias in the denominator of Cramér's V that causes tables with a smaller minimum dimension to yield higher levels of association. MCS provides a true baseline for each reported result so that any value of Cramér's V higher than that baseline represents association above random variation.

The same theoretical assumptions require us to follow necessary principles to conservatively interpret those results. Two principles are noteworthy.

1. Summary statistics from tables that have very close to the same dimensionality (same number of categories) can be compared to one another, especially if (a) they share one dimension in common, (b) all tables so compared are significant, and (c) they display a large enough N to yield relatively tight confidence intervals. For our purpose this means that we can compare the relative association of genre and authorship theories *within* a given language measure because we meet requirements a, b, and c.
2. Unless the data undergoes further norming,¹⁷ comparing the relative magnitude of association between different language measures (e.g. between lemmas and semantic domains) is contraindicated because forms with more categories (and hence fewer instances per category) have more potential for association in general¹⁸ and are drawn from a different distribution. Moreover, as the numbers of row categories (language

¹⁶ The Monte Carlo simulations were performed by randomly assigning the books of the GNT into groups of the same sizes as our theories of genre and authorship (8, 9, 10, 13 and 16). One hundred Monte Carlo simulations were performed for each linguistic measure by each theory tested by each bin size.

¹⁷ There are two methods that better compare results that differ markedly in category depth: calculating measures of association using standard deviates rather than raw counts, and calculating IQV (the Index of Quantitative Variation), which adjusts for the number of categories. See Reynolds, *Analysis of Nominal Data*, 62.

¹⁸ Per Reynolds, "Potential variation increases as the number of categories increases" (Reynolds, *Analysis of Nominal Data*, 61).

forms within a given language measure) inflates,¹⁹ the interpretation of association-based statistics becomes increasingly vacuous; Cramér's V becomes unreliable and PRE-based measures (G&K's tau and UC) converge to perfect but meaningless association.²⁰ To characterize meaningful association from meaningless association (i.e. category inflation) I have qualitatively compared the profiles of all language measures using the various binning and partitioning strategies and quantitatively calculated significance based on bootstrapping.²¹

Collectively, these considerations defined an initial workflow in which 605 tables were generated: eleven language forms by five bin categories (all, 25, 50, 100 and 200) by the eleven grouping theories (by each book, six authorship theories, and four genre theories). In addition, I ran this same design by an alternative bin strategy: the top 25, 50, 100, 200 and 400 most variant forms.²² This yielded a total of 1,210 total contingency tables, a representative subset of which I summarize in Tables 7-12. Other minor yet important steps were necessary to overcome analysis software limitations and other considerations.²³ The complete association dataset (Cramér's V,

¹⁹ Four linguistic measures display marked distributional inhomogeneity. Over 80% (4,185 of 5,314) of the individual lemmas in the GNT occur less than 10 times, 92% (16,336 out of 17,736) of the inflected lexemes (words) occur less than 10 times. Even more remarkably, 84.34% and 84.41% of the categories of proximity-derived and standard-derived clause complexes respectively contain only *one instance* in a category. For these tables, unbiased statistical estimates (those derived via bootstrapping) are especially indicated.

²⁰ In this data set, for instance, the lowest three deciles of categories from proximity-derived clause complexes contain only one form per category. In these cases, no distribution by grouping exists. Hence knowledge of the row category perfectly predicts the location of that measure in the column. This yields perfect association but is meaningless interpretively.

²¹ The primary finding we discovered was that the highest average association occurs in linguistic measures with the highest number of categories (tables with the lowest expected frequency per cell). This association may be caused by two opposite effects, one legitimate and one spurious. First and legitimately, lower frequency forms may display more patterned distributions (e.g. by author or by genre). This is an argument typically assumed in New Testament stylistic studies but also one that has never been thoroughly demonstrated using the whole distribution. Second, a component of this higher association may be spurious since tables with a large percentage of cells with low expected counts per cell violate the assumptions of nonparametric Fisherian statistical analysis. More extensive bootstrapping and the use of exact tests to characterize partitioned portions of the sparser lower parts of the contingency table are in work to more fully characterize this error component.

²² Adjusted standardized residual was used as before.

²³ The IBM-PASW (SPSS) crosstab procedure allows a maximum of 1,000 rows or columns. Consequently, using SPSS to calculate Cramér's V and Goodman and Kruskal's tau was not possible for all 5,413 lemmas, all 17,736

G&K's tau, and UC by bin by language measure) can be inspected via the various permalinks we include in the body of this work.

6.3 Testing Bivariate Association, Direction and Significance

It should be noted that from this point forward (through chapter nine) I will follow the experimental sequence defined in Section 4.4.2. That is, (a) an objective will be stated (or clearly implied by the title), (b) the language measures and groupings to be used in the analysis will be defined, (c) the analysis method will be detailed and (d) optionally any requisite interpretive frameworks necessary to understand the data will be stated. Per the title above, our first quarry will be to test the association, direction and significance between the six authorship theories and the four genre theories defined in Section 5.3 and 5.4.3.

IMPORTANT NOTE: Because evaluating quantitative findings is foreign to many, two steps will be taken to demystify both the terminology and the findings here. *First*, while findings must be stated in an appropriate technical register, when it seems that the terminology is unclear, the therefore sign (\therefore) will be used and a *very brief* explanation will follow in a nontechnical register. *Second*, where needed a hermeneutical interaction section will be added to the discussion section. If that section is present, an imaginary interlocutor will ask nontechnical *interpretive* questions of the data. This section will proceed in Q&A format.

Statement of Objective: A) Test the overall *strength* (association) and statistical significance of the six authorship and four genre theories with each of eleven representative language measures in the GNT. B) Test the overall *direction* of the association between the language measures and authorship and genre. C) Test for *significance* in the differences between the theories.

distinct lexemes (words), all 1,412 clauses, and all proximity-based and standard-based clause complexes (11,263 and 12,304) respectively. To address this limitation, we implemented two steps. First, we rebinned (combined) variables in different configurations. Second, we coded our own cross tabulation procedure without a row or column limitation and calculated Cramér's V and Goodman's tau up to and including the unbinned case where no low frequency forms are combined with any others.

Language Measures Used: Lemmas (the canonical lexeme), inflected lexemes (“words”), semantic domains, semantic subdomains, three traditional grammatical annotations of the GNT, word groups, clauses, and the two types of clause complexes.

Statistical Test(s) Used: Association was measured with Cramér’s V and G&K’s tau, and unbiased bootstrapped estimates of Cramér’s V and G&K’s tau. Confidence intervals (CI) around Cramér’s V and G&K’s tau were also calculated via bootstrapping. Directionality was tested via G&K’s tau using the theory in question (authorship or genre) as the dependent variable.

Interpreting Tables 6.1-6.6: The following interpretive points should be understood.

1. The tables that follow juxtapose Cramér’s V and G&K’s tau for the purposes of comparison.
2. The titles under G&K’s tau reflect that variously it is referred to as a measure of *association and directionality* whereas the titles under Cramér’s V indicate (properly) that it is a measure of *association*.
3. The error bars above the column represent the unbiased estimates of G&K’s tau and its 95% CI (confidence interval) calculated using bootstrapping.
4. The units of Cramér’s V or G&K’s tau as the case may be, are projected on the right axis.
5. The columns in all charts follow the same order from left to right: the control group (ungrouped books), the six authorship theories, and the four genre theories (the lighter columns).
6. The leftmost darker column is the original Baur hypothesis, and the rightmost darker column is the mainstream critical view of authorship.

7. To avoid sparse tables (and consequent violations of analysis assumptions) for language measures with large numbers of categories, these categories are reported using binned data.
8. Binned data means that the top most frequent categories are used. Thus “25 bin” in the leftmost chart in Table 6.1, means that the table had 26 categories: the Top 25 lemmas by frequency and a final category that contained all the other lemmas with a frequency less than 25. Last, in some cases the Top N categories in the data charts (not shown but at the [permalink](#)) were by variance (by AAVASR) rather than frequency.

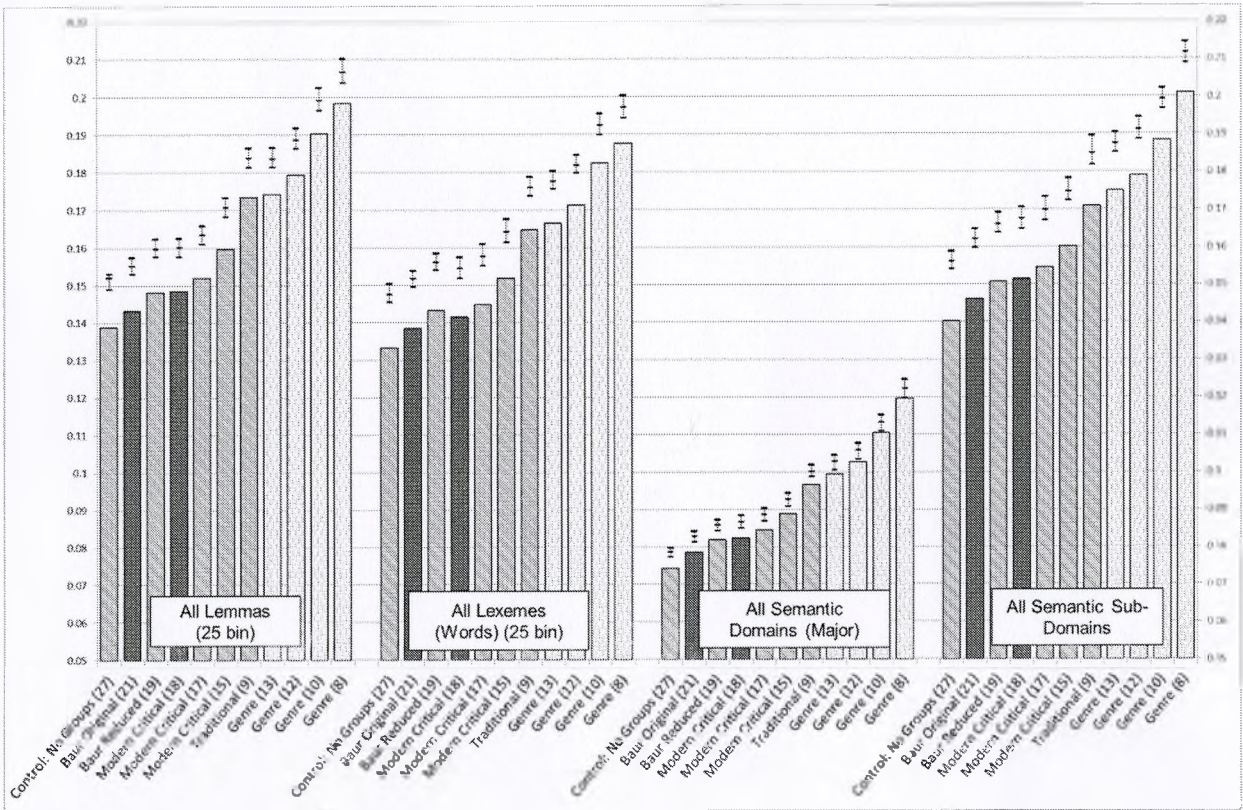


Table 6.1 Association at the Rank of Lexis: Cramér's V

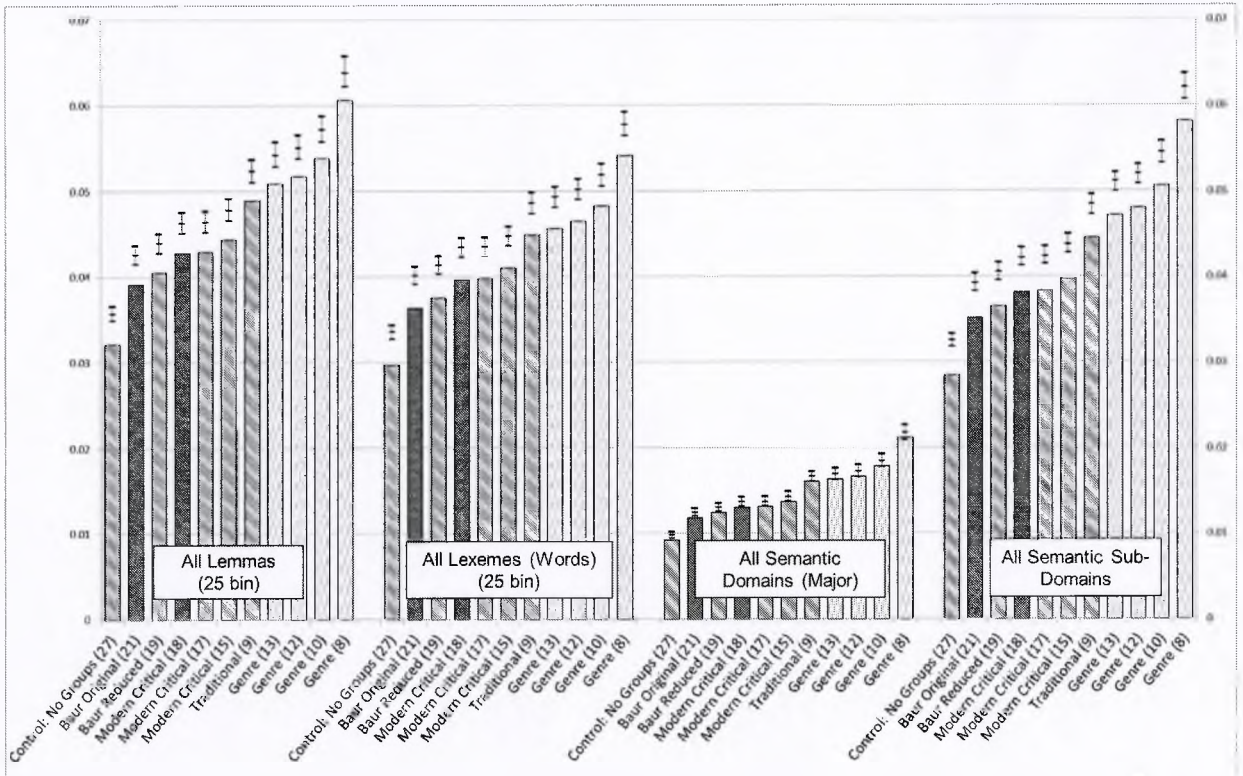


Table 6.2. Association/Directionality at the Rank of Lexis: G&K's tau

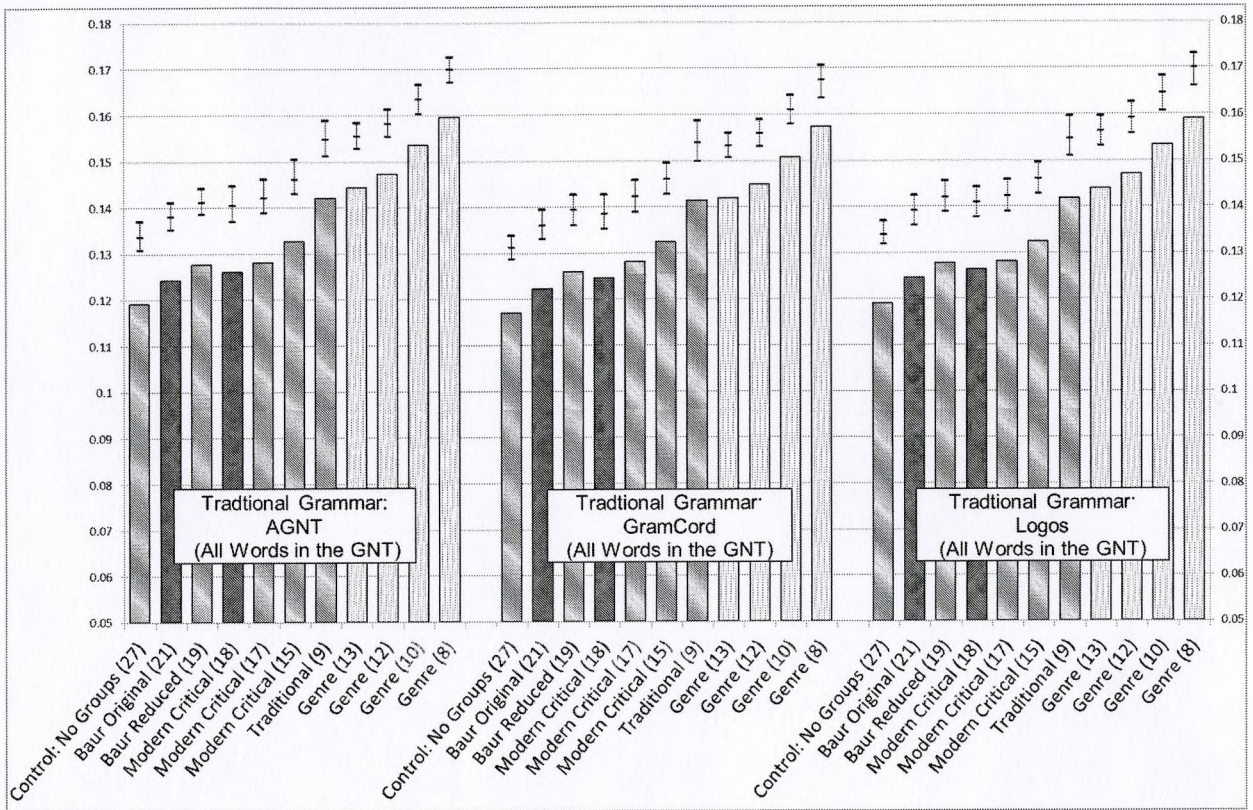


Table 6.3 Association Using Traditional Grammar: Cramér's V

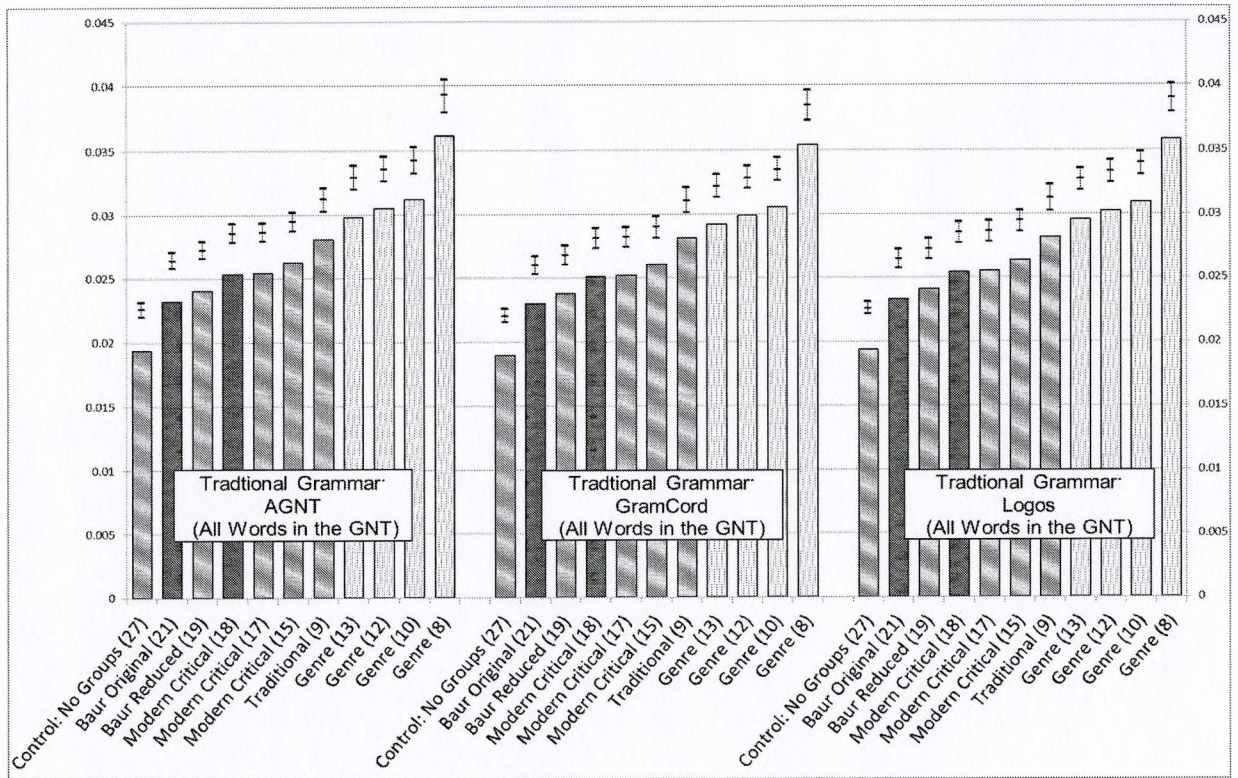


Table 6.4. Association/Directionality Using Traditional Grammar: G&K's tau

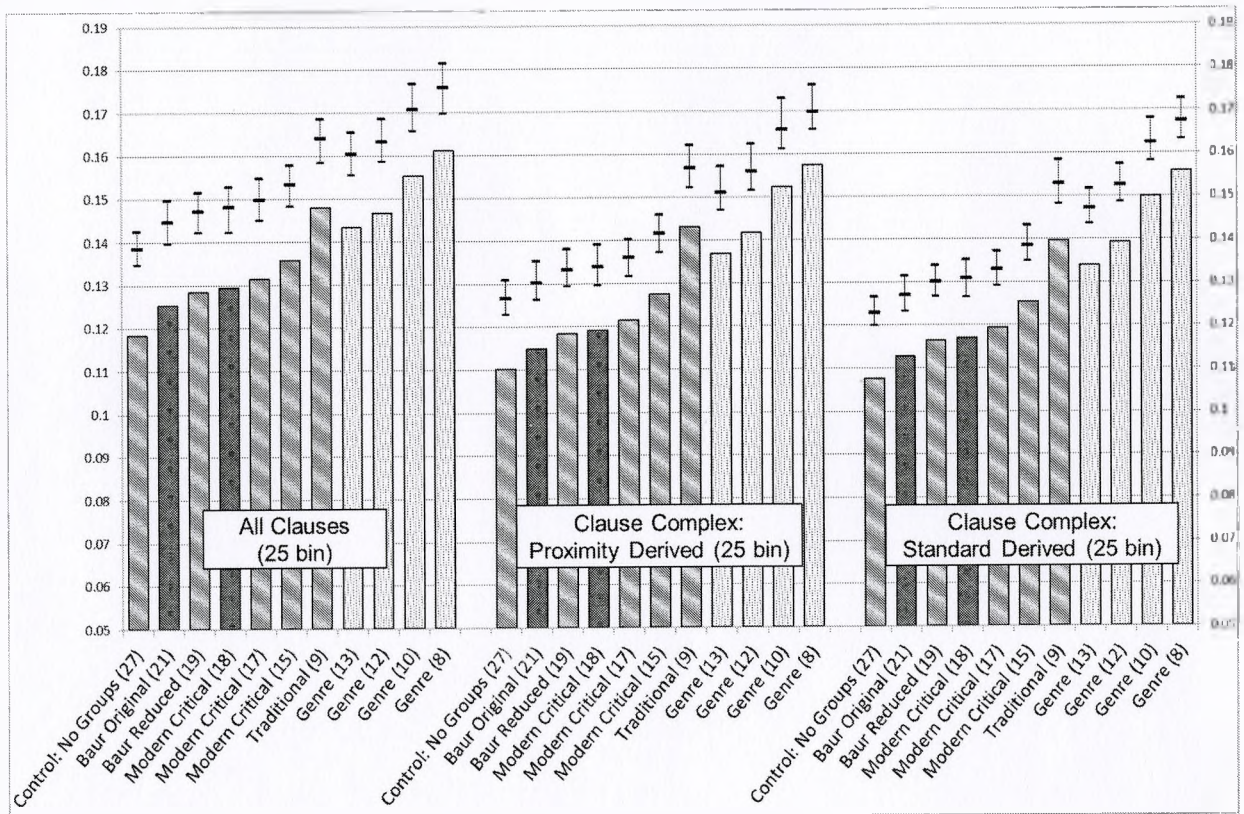


Table 6.5. Association Above the Rank of Lexis: Cramér's V

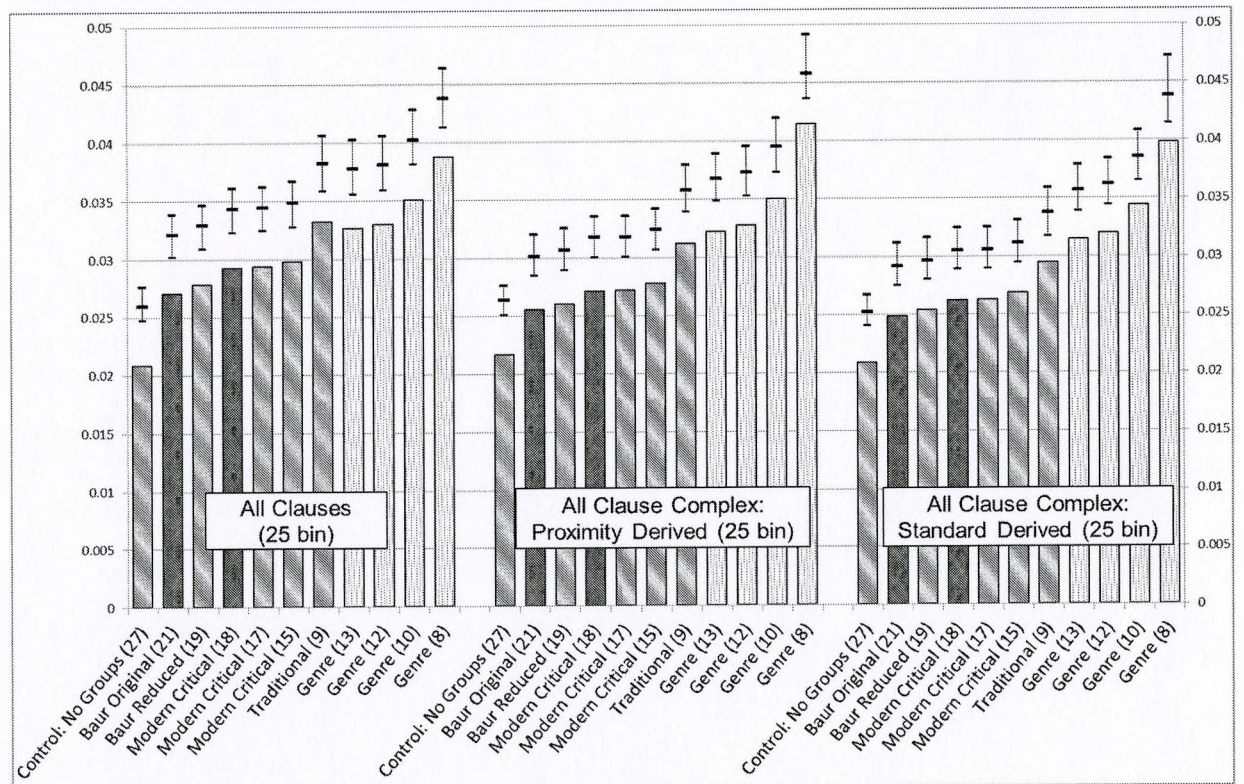


Table 6.6. Association/Directionality Above the Rank of Lexis: G&K's tau

6.3.1 Findings²⁴

- All contingency tables are highly significant ($p > 0.000$) using table level chi-square.
∴ For each table (each bar in each chart represents one table) there is significant association between the rows (the language measure) and the columns (the authorship or genre theory).
- The data profile seen in tables 6.1-6.6 retained their characteristic shape whether one looks at all language forms or bins.²⁵ (For other bins see the permalink.)
∴ The highest association between language forms among authorship was found with the nine-group (traditional theory). The highest among genre was genre eight theory. This occurred regardless of whether the measure was Cramér's V (6.1, 6.3 or 6.5) or G&K's tau.
- Word groups displayed a distribution so unimodal (70.6% are a single head term) and showed such low association with the language measures that further analysis of word groups is contraindicated for the time being. This leaves ten language measures to be tested.
∴ 70.6 percent of the word groups had only one element in them. This made the table-level significance so low that we dropped word groups from further analysis.

- ***High Similarity Between the Data Profiles of Cramér's V and G&K Tau:***

Perhaps the second most striking finding (after the shape of the profiles discussed above) is the high similarity seen between the Cramér's V profile and the G&K's tau profile. Due to this, all the findings discussed below will relate to G&K's tau. Any differences between G&K's tau and Cramér's V will be reported in the footnotes.

- ***Higher Association of the Authorship or Genre Theories Compared to the Control Group:***

All theories of authorship and genre were significantly more associated with all ten language measures (95% CI) than the control in which no texts are grouped with others.²⁶
∴ The differences between the categories within each of the ten theories is greater than the differences between the categories of the control groups (in which each category is a separate book in the NT). Thus each theory displays more association with that given language measure than the control group.

²⁴ In Section 6.3 the CI for Cramér's V was calculated via percentile bootstrapping (100 samples) and significance was determined if there was no overlap between the intervals.

²⁵ In addition to the above data, the four linguistic measures with the largest number of categories (lemmas, words, and clause complexes) were binned at even smaller bin combinations (e.g. 20, 10 and 5). The data profile above still held in every instance. The only case in which authorship theories slightly exceeded genre was with clause complexes when all low-frequency forms remained uncombined and then only with Cramér's V (not G&K's tau).

²⁶ For Cramér's V : All theories of authorship with less than 18 authors were significantly different (95% CI) from the control set for all ten linguistic measures. Moreover, for eight out of ten linguistic measures (all except semantic domains) the original Baur hypothesis was not significantly different than the control set.

- **Higher Association of Genre Over Authorship:**

- All critical theories of authorship are less associated with all language measures tested than any theory of genre. For nine of ten measures that difference was significant.²⁷
∴ *This reflects the first clear bivariate (language measure by theory) indication that genre as a class of theories is more associated with the language of the GNT than authorship.*
- The statistical difference between traditional authorship theory and the various genre theories is less than the statistical difference between the critical theories and genre.²⁸
∴ *Traditional authorship seems to be more “genre-like” than “authorship-like.”*

- **Higher Association of Traditional Authorship Compared to Other Authorship Theories:**

Traditional authorship displayed higher association with all language measures than any other authorship theory. For six of ten measures (ten of ten with Cramér’s V) that difference was statistically significant.

∴ *If the question is, “What is the authorship theory whose categories are the most different from one another?” the answer is, “Traditional authorship.”*

- **Higher Association of Genre Eight Theory Compared to Other Genre Theories:**

The eight-category genre theory was significantly more highly associated with all ten tested language measures than the thirteen- and twelve-category genre theories.²⁹ The eight-category genre grouping is even more dominant with G&K’s tau being significantly higher than all other genre theories except at the rank of the clause.

∴ *That is, if the question is, “What is the genre theory whose categories are the most different from one another?” the answer is, “Genre 8.”*

- **High Directionality:**

The data demonstrate by G&K’s tau that language predicts the theory (either authorship or genre) on average fifteen times more than the converse. (Data not shown.)

∴ *As a proportional reduction of error metric, G&K tau can be “run” in two ways. That is, the statistic can calculate whether the rows (the language measure) are a better “explanation” for the columns (the theory whether authorship or genre) or the converse. The former is confirmed as one would expect.*

²⁷ For Cramér’s V: All theories of genre were significantly more associated (95% CI) with nine of the ten linguistic measures than all critical theories of authorship.

²⁸ Specifically, traditional authorship is significantly different from the genre theories 13, 12, 10, and 8 for one, four, eight, and ten linguistic measures respectively. For Cramér’s V: Traditional authorship is significantly different from the 13, 12, 10, and 8 category genre theories for zero, one, six and ten linguistic measures respectively.

²⁹ For Cramér’s V: The 8-category genre theory was significantly more highly associated with all ten tested linguistic measures than the genre 12 and 13 theories. The 8 and 10 category genre theory was significantly more highly associated with five of the ten tested linguistic measures compared to the genre 12 and genre 13 theories.

6.3.2 Discussion of Findings and Hermeneutical Interaction

Five findings, in particular should be discussed in greater detail.

1. Genre is more highly associated with eight of ten language measures than any theory of authorship by G&K's tau. *This fact strongly implies that genre is a larger component than authorship in the total observed variation in the GNT.* We will further test this inference via decompositional analysis in Chapter Seven.

Q: Since genre theories are more highly associated with language than authorship, isn't our job done? Haven't we proved that genre comprises a larger percent of the total variation in the GNT?

A: Actually no. Association is not the same thing as "surgically removing" the variation caused by authorship and genre and weighing them side by side. All we have done so far is group the texts of the NT into authorship and genre groupings and look at their association by different "cuts" of linguistic data. We have not yet formally separated authorship from genre. That is the task of multivariate analysis (Chapter Seven). Said differently, all the variation is still "inside" each of the bars in each of the charts. This exercise did, however, demonstrate that, in toto, genre groupings are more associated with all language measures than authorship. It would be difficult for this to be the case if genre was NOT a larger percent of the total variation than authorship. However, because we do not yet know whatever else may cause variation in the GNT (e.g. the various kinds of sociolectic and dialectic variation for instance) and whether those other causes covary more with genre or authorship—we cannot yet make that equivalence.

2. The consistently higher Cramér's V and G&K's tau for traditional authorship compared to all other critical views of authorship was both unexpected and intriguing. To further confirm this finding requires that we (1) inspect other measures of association (e.g. the index of diversity and qualitative variation and the like),³⁰ (2) decompose (separate) authorship effects from genre effects and (3) inspect measures

³⁰ For the derivation and utility of these indices, see Reynolds, *Analysis of Nominal Data*, 61–64.

other than association.³¹ In this study we pursue steps two and three in Chapters

Seven and Eight of this work. We will reserve step one for another study.

Q: Why is the traditional authorship theory more highly associated with all the language measures than any other authorship theory?

A: Why this is the case is not yet clear. This could, for instance, be because the nine authors fall into categories that are similar to the genre categories. This would occur if authors in the nine-group theory better mapped to fundamental genre categories such as Aune's four categories of genre. More research is indicated.

3. The significantly higher association of the eight-category genre grouping over the ten-category genre grouping for some language measures must be further explored.
4. The overall measure of association for G&K tau is low. Mitigating this concern, however, are three facts: (1) The level of association is highly significant ($P < .001$ in all cases reported here). (2) The CI's are tight and reproducible. (3) When G&K's tau is used to analyze the low-frequency data partition (the lowest quartile of categories by frequency) G&K's tau is much higher (with values above 0.35) and yet displays the same profile qualitatively.³² To definitively determine whether this level of association was above random chance, we coded a Monte Carlo simulator. The Monte Carlo runs demonstrated that for both Cramér's V and G& K's tau all reported statistics are far above the Monte Carlo baseline.³³
5. Tables 6.1-6.6 demonstrate that, generally, G&K's tau increases as the number of groups decrease. One might reasonably surmise that the higher association seen with genre is an

³¹ The article by Preuss and Vorkauf, "The Knowledge Content of Statistical Data," 133–61 and the chapter by Yao Karmeshu, ed., *Entropy Measures*, 115–36 are especially helpful in this regard.

³² The raw data can be inspected at:

ftp://decisionsupportsciences/stat_ling/Lingusitic_Measure_by_Frequency_Deciles.xlsx.

³³ In our Monte Carlo simulations the NT books were randomly binned into 8, 9, 10, 12, 13, 15, 17, 18, 19, and 21 categories – the same size groupings as our various authorship and genre theories. Then 100 crosstabs sets were run for each of the nine linguistic measures by each of these category groupings. The G&K's tau in every case was significantly above random chance. See the Monte Carlo tab within ftp://decisionsupportsciences/stat_ling/Association_By_Book_4Genre_6Authorship_Functional_Bootstrap.xlsx.

artifact – because genre is measured in fewer overall groupings than authorship. This observation is inaccurate for four reasons:

- a. The data itself demonstrates that the thirteen-, twelve-, and ten-genre categories have higher association with the various language measures than the nine-category authorship theory.
- b. The tables that comprise each profile are, in actuality, very close in degrees of freedom. That is, it is only the theory grouping dimension (the columns) that differs between the bars within a profile.
- c. Such a construal misunderstands the mathematics of both Cramér's V and G&K's tau. Cramér's V adjusts for the size of the smallest dimension, which, in our data *is* the group size.³⁴ Moreover, G&K tau is a PRE measure, and hence is dimensionally invariant.
- d. Empirically, when authorship and genre theories are constructed with fewer categories than eight or nine there reaches a point where association *decreases*.³⁵

³⁴ In addition, Monte Carlo simulations adjust for the well-understood (but slight) bias in the denominator of Cramér's V that causes grouping with few categories to yield higher levels of association.

ftp://decisionsupportsciences/stat_ling/Association_By_Book_4Genre_6Authorship_Functional_Bootstrap.xlsx

³⁵ These were created using the nine-author theory and eight-category genre theory respectively. A total of 5,000 separate random recombinations (Monte Carlo simulations) were run for each group size (21, 20, 19, 18, etc.) and this data demonstrated that past a certain point the association decreased.

6.4 Bivariate Information Theory: Another Lens for Viewing Authorship and Genre

Chapter Nine of this study will present Information Theory (IT) as a separate, complementary mathematical “lens” to further understand the relationship between authorship and genre in the GNT. One of the central measures presented there will be the Uncertainty Coefficient (UC). While the UC is often erroneously represented as a measure of association, mathematically it is actually not a measure of association *per se*, but rather is *the normalized total amount of information sent from a sender that is unambiguously received by a receiver*. As such, it tells us, in a normed way, which theory of authorship or genre communicates the most information, and which theory obscures the most information. Moreover, given the assumption of pseudepigraphal theory that low-frequency information such as *hapax legomena* is more associated with authorship, and given that such low-frequency information is more “marked” (it counts for more) in information theory,³⁶ exploring information theory measures in Chapter Nine is wholly appropriate. Since we have dedicated Chapter Nine to information theory, why should we preview UC information here? There are two reasons. *First*, UC is a bivariate measure, and this chapter is dedicated to exploratory bivariate.³⁷ *Second*, it is very seldom recognized that the mathematics of hierarchical log-linear analysis (HLLA) to be presented in the next chapter are closely related to information theory (IT).³⁸ Presenting an instance of IT data here will serve as a

³⁶ While it is considered axiomatic that low-frequency forms communicate more information in natural language than high-frequency words, it oversimplifies the case to limit information content only to relative frequency. Information content is also markedly affected by rare words used frequently and topicality; Finn, “Word Frequency, Information Theory, and Cloze Performance,” 508–35.

³⁷ We will disentangle the effects of authorship and genre in chapter seven; reviewing information theory-based data here informs our experimental design for that section.

³⁸ See Goodman, “A General Model for the Analysis of Surveys,” 1050; Bishop et al., *Discrete Multivariate Analysis*, 344–48; Krippendorff, *Information Theory*, 89–90.

preface of sorts to Chapter Seven. Accordingly, a preview of those findings is presented in Table 6.7 below

6.4.1 Bivariate Information Theory Findings

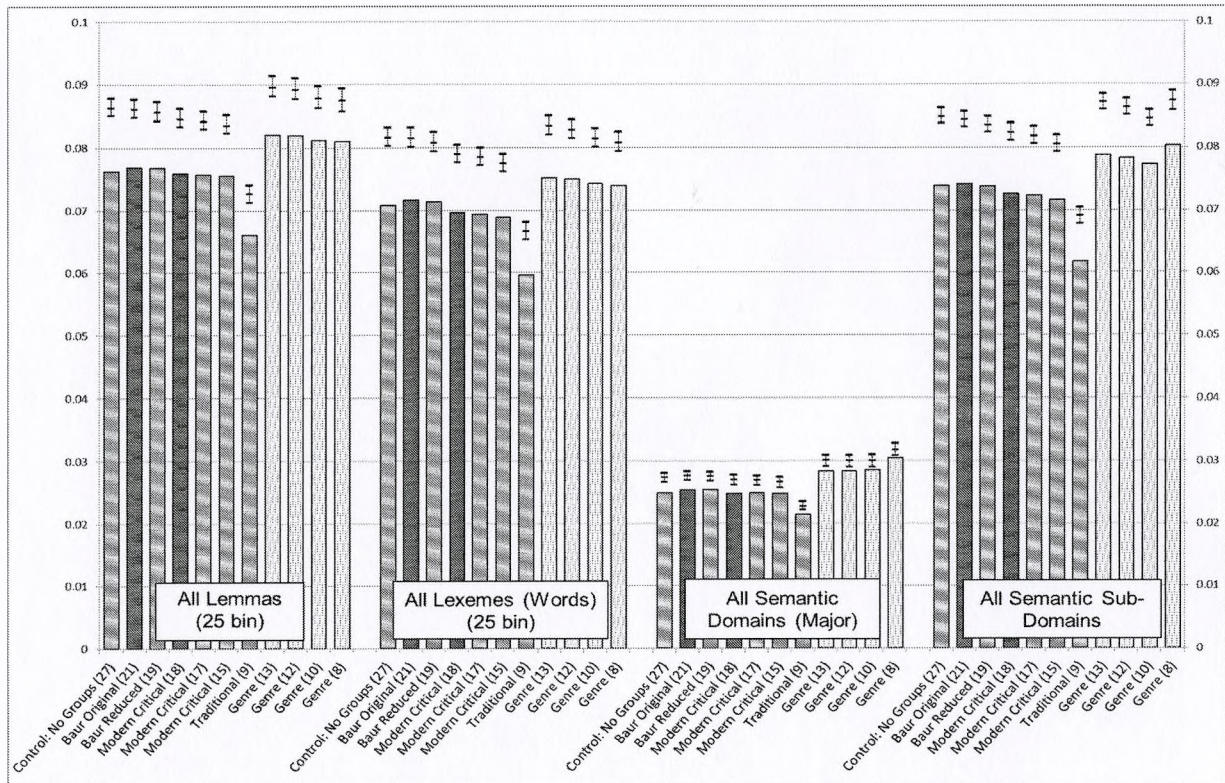


Table 6.7· The Uncertainty Coefficient (UC) at the Level of Lexis (Single Words)

6.4.2 Discussion of Findings

Inspection of Table 6.7 reveals three facts relevant to genre and authorship. First, the shape of the UC profile is NOT qualitatively similar to either Cramér’s V or G&K’s tau. This is quite expected because the “lens” of UC does not represent association but *normed total information*. A deeper explanation of this and its importance will be presented in Chapter Nine. Second, as before, genre generally communicates higher normed information than authorship, just as genre was more highly associated than authorship via Cramér’s V and G&K’s tau (per Section 5.4). Third, and most strikingly, whereas almost uniformly the traditional nine-author theory shows the *highest level of association* (via Cramér’s V and G&K’s tau), here it

communicates *the lowest level of normed information* (UC). This differential obscuring of information is no artifact. It occurs with every bin of the data and for every language measure.³⁹ How can an authorship theory that has the *highest* level of association transmit the *least* normed amount of information? Moreover how does all this fit with the low-frequency assumption of the difference between genre and authorship in the GNT, or for that matter the issue pseudepigraphy? This I discuss below.

6.4.3 Explaining the Marked Destruction of Information with Traditional Authorship

As this work has stressed, the error-based measures (Cramér's V and G&K's tau) and entropy-based measures⁴⁰ (UC) provide distinctly different perspectives into the same data distribution. Said differently, they "look" at different aspects in the data, and hence provide a complementary picture of what is going on in the language of the NT. Yet it adds nothing to our understanding of the language of the NT to merely issue a *Panglossian* fiat that this is so. Rather, we must demonstrate what fraction, partition or other aspect of the distribution causes the association of the traditional authorship theory to be *highest* in error-based association (Cramér's V and G&K's tau) among the authorship theories while being *lowest* in the entropy-based uncertainty coefficient (UC). Furthermore, we must explain why the other authorship theories remain relatively unaffected. Said less mathematically, the nine-category authorship theory *for some reason* destroys or loses more language information (recall that the UC communicates information) than any other authorship or genre theory. Our task, if possible in this initial exploration, is to find out why.

³⁹ It also occurs in partitions of the data, where high-frequency data is separated from low-frequency data by decile.

⁴⁰ *Entropy-based* measures derive from Shannon's theory of entropy, and lie at the core of information theory; Shannon and Weaver, *The Mathematical Theory of Communication*. We provide a primer on these measures in Section 9.2.

One proposal is that the *characteristic information* of the language of the NT may be explained far better by traditional authorship (Cramér's V and G&K's tau) but the *unique information* is explained far worse. We might also surmise that the characteristic information can be safely equated with high-frequency information (that's what makes it characteristic)⁴¹ and unique information, whether it be unique to a genre or author, *tends* to be low-frequency information.⁴² To explain how characteristic information can be so well explained by traditional authorship while unique information is so poorly explained, let's return to the two assumptions of pseudepigraphy. The first is a very reasonable one, that there are unique patterns to an author's style. With this scarcely anyone can disagree. A second and often unstated assumption, however, is far more problematic, namely, that that total patterning in the GNT is due wholly or even predominantly to authorship. Given that the UC findings just presented confirm that a higher amount of information across all language measures is left unexplained by the traditional authorship theory, we can now ask why. *Assuming the most parsimonious case in which genre and authorship are the main covariates which can best explain the language patterning in the GNT, and that other context of situation, context of culture or other exogenous effects are*

⁴¹ High-frequency information may exist for any number of reasons. It may be demanded by the genre, such as the predominance of the perfective verbal aspect (aorist tense) as the standard narrative tense form (Mathewson, *Verbal Aspect in the Book of Revelation*, 50). Alternatively, it may be a discourse device by a given author whose style demands lexical repetition to create cohesion in the text; Halliday and Hasan, *Cohesion in English*, 278–81.

⁴² Of course, an author can characteristically repeat a phrase unique to him or her many times in a corpus or even in one of her or his books. This would result in unique information being found, quite possibly, in high frequency for that author. But this seldom occurs. Markedness theory, which is closely allied to information theory, typically argues that only low-frequency information is marked. This misunderstands the cognitive basis of perception. It is my contention that the brain differentially notices *both* low- and high-frequency information because they stand out from the expected frequency.

comparatively negligible,⁴³ the traditional authorship theory can leave more information unexplained for only one of two reasons:⁴⁴

- **Reason #1:** Different *authors* are being grouped together to form the traditional nine-author grouping, thus destroying unique authorship information. This is the consensus stance of pseudepigraphal theories. According to this view we assume (as pseudepigraphal theories do) that the measured unique (largely low-frequency) information is disproportionately *authorial*.⁴⁵ Accordingly, as we look at lower and lower frequency fractions of language, the magnitude of the effect of authorship compared to genre should *increase*. In this scenario the nine-author theory would be *expected* to obscure more low-frequency authorial information, since multiple *authors* are being shoehorned into one author.
- **Reason #2:** Different *genres* are being grouped together (per scholars supporting traditional authorship) thus destroying unique *genre* information. Accordingly, as we look at lower and lower frequency fractions of language, the magnitude of the effect of genre compared to authorship should *increase*. In this case the unique information in the traditional authorship theory is being obscured for another reason entirely, because multiple authors (such as Paul) can and do write across multiple genres. What is being destroyed, therefore, per the traditional nine-author view, is disproportionately context-specific *genre information* because distinctly different *genres* are being shoehorned into one author.

Given these two very different explanations for the same phenomena, if we can devise an experimental design that demonstrates that the preponderance of the unique information (found disproportionately in the lower frequency partitions of language in the GNT) is associated with *authorship*, then such a design will lend credence to the theory of pseudepigraphy as it has been traditionally argued. Conversely, if we can devise an experimental design that demonstrates that the preponderance of the unique information in the GNT is associated with *genre*, then that

⁴³ It is to be understood that I am not asserting that other effects do not exist. In sociological research, in fact, such effects may be sizable. Moreover, in modeling sociocultural causation, effects are quite often multivariate and multilayered. (Multivariate means that two or more factors contribute to a given effect, and multilayered means that observed factors contribute to unobserved latent factors, which, in turn, contribute to the observed effect.) Multifactorial, multilayered models are built and estimated using latent class modeling (Long, *Covariance Structure Models*). The objectives of this initial experimental design here, however, are simpler, driven by two factors. First my thesis is to explore the relation between genre and authorship. Second, the implicit assumption is that genre and authorship are the main covariates and other confounding effects are relatively negligible.

⁴⁴ We have already mentioned a third – that grouping texts together naturally destroys information.

⁴⁵ Said more accurately, for this interpretation to be affirmed of the total joint distribution, we must assume that the preponderance of the unique information in the GNT is associated with authorship. To unambiguously make this conclusion, however, requires decomposing the effect of authorship from genre. This is the work of Chapter Seven.

design would support the traditional nine-author theory.⁴⁶ This we do in Chapters Eight and Nine.

6.4.4 Using the Hypothesis of Differential Information Destruction Seen in the Traditional Nine-Author Theory to Inform the Experimental Design in Chapter Seven

How then can we determine which of the two possibilities just expressed in Section 6.4.3 best explains the differential loss of information seen in the nine-author theory? Procedurally, designing such an experiment is actually quite straightforward. It requires only five tasks:

1. *Mathematically disentangle genre from authorship.* Decompose the saturated variance structure of a three-way (language measure by authorship theory by genre theory) contingency table using HLLA. This yields two, two-way interaction effects of immediate interest: (1) the interaction term between the language measure with authorship, which by definition is authorship controlled for genre, henceforth (A|G)⁴⁷ and (2) the interaction term between the language measure with genre, which by definition is genre controlled for authorship, henceforth (G|A).⁴⁸

⁴⁶ An important question should be raised at this point. If an association with the unique low-frequency information in the GNT can be established with either authorship or genre, what can we say about the characteristic high-frequency information? It is tempting to say that if the low-frequency information is associated with, say, authorship, then the high-frequency (characteristic) information must be associated with genre. Unfortunately, this line of reasoning does not *necessarily* follow. Simply because it is established that low-frequency information is *predominantly* associated with a given effect (in this case an interaction effect) it does not necessarily follow that high-frequency information is predominantly associated with another effect. There may, for instance be five or ten effects at work here! Or a given effect may be expressed *homogeneously* with respect to the data partitions. Even if we cannot directly infer the distributional state of one effect (genre) when the other (authorship) is known, we can propose the following testable hypothesis. If, by definition, the low-frequency fraction of language is being predominantly transduced with *entropy-based measures*, might the difference in profiles be because the high-frequency characteristic information of the text is being predominantly transduced with *error-based measures*? Moreover, depending on what the experiment finds, it may indeed be the case that the profiles seen with the traditional nine-author theory in the NT is due to the fact that association-based measures differentially focus upon the higher frequency characteristic information of each author (or genre) while information-based measures differentially focus upon low-frequency genre (or author) information. The point of this extended exercise in inference is that this kind of thinking leads to following up with other experimental designs.

⁴⁷ A|G is the two-way interaction between a given linguistic measure and a given authorship theory.

⁴⁸ G|A is the two-way interaction term between a given linguistic measure and a given genre theory.

2. *Compare authorship (A|G) to genre (G|A) “head to head” for every possible combination of theories.* Said in terms of the title of this work, (A|G) is the effect of authorship with language disentangled from genre, and (G|A) is the effect of genre with language disentangled from authorship.
3. *Test if low-frequency information is more associated with authorship (A|G) or genre (G|A).* Exploring low-frequency forms can be achieved two ways, by binning high-frequency forms (HFF),⁴⁹ or by removing them outright (partitioning), thus leaving only low-frequency forms (LFF).⁵⁰ Then rerun steps 1 and 2. If low-frequency information is more associated with authorship than genre (per the critical/pseudepigraphal hypothesis), then binning HFF or inspecting only the LFF partition will strengthen authorship (A|G) compared to the HFF partition. Conversely, if low-frequency information is more associated with genre than authorship (per the traditional authorship hypothesis) then inspecting the LFF partition will strengthen genre (G|A) compared to the HFF partition.
4. *Compare the competing authorship theories (A|G) with one another.* The authorship theory (A|G) that explains the most variation in the text has, *ceteris paribus*, the highest likelihood of being the primordial theory of authorship.
5. *Compare the competing genre theories (G|A) with one another.* The genre theory (G|A) that explains the most information has, *ceteris paribus*, the highest likelihood of being the primordial theory of genre.

These five steps will occupy our efforts in Chapter Seven.

⁴⁹ Binning retains all forms, but obscures their individual contributions by combining them into a final category.

⁵⁰ Partitioning, in contrast, eliminates forms, but no frequency categories are obscured by binning.

7 Results Part II: Disentangling GNT Authorship and Genre Using Multivariate Analysis

7.1 Disentangling Authorship from Genre Using HLLA: A Primer

7.1.1 HLLA Theory

Hierarchical Log Linear Analysis (HLLA) models are specialized mathematical models that explicitly leverage the insight that the total variance in any contingency table must necessarily be the sum of the variance of each dimension of the table (its main effects) plus the total possible interactions between those dimensions.¹ For example, for a table with three dimensions (e.g. language measure by genre by authorship) there are necessarily seven components which explain the total variance in the table. These seven components are: the variance for each of the main dimensions of the table (its *main effects*) taken separately (i.e. language measure, genre, and authorship), three two-way interactions (i.e. language measure by authorship, language measure by genre, and genre by authorship) and one three-way interaction. These seven components constitute a *saturated* HLLA model in which there is no error between the observed values and the prediction of the model because all the main effects and interactions are quantified. As HLLA is most typically used, the model calculates the strength of all its terms (in our case the three main effects plus the four interaction effects) to discover which one(s) can be removed and still yield an acceptably small amount of total error (residual) between the observed and predicted values in the cells of the contingency table. The removal of the least important effects is consistent with traditional modeling praxis which seeks to explain the observed frequencies most parsimoniously by using the fewest possible terms. To develop the most parsimonious model, HLLA typically employs a type of stepwise calculation and

¹ For an introduction to the method see Knoke and Burke, *Log-Linear Models*, 11–29.

verification termed *backward elimination*. The practitioner is then free to inspect the HLLA model after each step and determine, assisted by a set of associated statistics, at which point the most parsimonious model has been achieved.

Given this description, two important capabilities of HLLA modeling can be immediately understood as relevant to our task of disentangling genre from authorship in the GNT. First, because HLLA models have the ability to decompose the total variance into its constituent components we have at our disposal a quite remarkable tool – one that gives us a method of separating or *disentangling* the effect of authorship from the effect of genre. For this reason, as defined before, the two-way interaction term “genre by lemma” actually means “the variance from the interaction between lemmas and genre when controlling for authorship” (G|A).² The second advantage to using HLLA modeling, however, is just as important as the first. Because HLLA removes the weakest terms first (where weakest is defined as the term that contributes least toward explaining the observed variation) it rank orders whether genre or authorship is “stronger” for a given model. Said more technically, if the removed term is, for instance, the interaction between the language measure and genre, and if its parallel term of the same *order*³ remains in the model (e.g. language measure by authorship) then we can clearly consider the first removed term to be less associated with the main effect that they share in common (i.e. the language measure). Last, because we will be interpreting so many HLLA models, we need a simple decision rule to help us determine for a given model whether authorship is “stronger” than genre or vice versa. To develop this decision rule, two pieces of information are necessary. First, as they are typically implemented, HLLA models produce *deletion parameter statistics*

² See DeMaris, *Logit Modeling*, 16.

³ In all decompositional mathematical models, terms that have the same number of dimensions measured in them are said to be of the same *order*. Hence (A|G) and (G|A) are terms of the same order.

which allow the researchers to actually assess how much explanatory information is lost when a given parameter is eliminated from the model. A variety of appropriate normed statistics are available to allow the two terms of interest, $A|G$ and $G|A$, to be compared to one another.⁴ The one most readily available from the deletion parameter statistics of the HLLA table, however, is chi-square divided by its associated degrees of freedom. This quotient, classically referred to as *reduced chi-square*,⁵ is equivalent to the “strength” of that term, the amount of normed information lost to the model when that term is removed.⁶ Second, armed with this statistic and the previously provided description of how backwards elimination works, the decision rule to determine which effect ($A|G$ or $G|A$) is “stronger” becomes readily apparent. A given interaction term, say $A|G$, is “stronger” than the other interaction term of the same order ($G|A$) when (1) the second term ($G|A$) is *either* removed in a step prior to the given term ($A|G$), *or* (2) in the case in which two terms ($A|G$) and ($G|A$) are removed at the same step, the term which expresses the larger *normed* variance (in our case reduced chi-square) is the “stronger” term. To avoid repeating this somewhat lengthy technical definition in the following pages, the term ($A|G$ or $G|A$) that meets these conditions will simply be referred to as the “stronger” term.

⁴ Any appropriate bivariate normed measure of categorical variance can be used, such as Cramér’s V. Reduced chi-square however is the more robust statistic than Cramér’s V overall in comparing $G|A$ to $A|G$. This is due to the fact that in those few cases in our design set in which the linguistic dimension (which is always present in $A|G$ and $G|A$) has *fewer* categories than the categories of authorship or genre, Cramér’s V fails to take into account row invariance—that the language dimension of $A|G$ and $G|A$ are identical.

⁵ Reduced chi-square is used in this way in the multivariate parametric parallel to HLLA, Multiple Analysis of Variance (MANOVA).

⁶ We have also calculated more conservative statistics in parallel with reduced chi-squared, Phi and the now familiar Cramér’s V..

7.1.2 HLLA Praxis: Exploring Pseudepigraphal Hypotheses after “Disentangling”⁷

Statement of Objective Three questions arise once HLLA has disentangled authorship effects from genre effects: (1) Is authorship (A|G) or genre (G|A) “stronger” in the GNT?⁸ (2) Is authorship (A|G) or genre (G|A) more associated with low frequency data? (3) How relatively “strong” are the authorship theories (A|G) to one another after being disentangled from genre?⁹

Language Measures. Nine of the original eleven language measures were selected for this exercise: lemmas (the canonical lexeme), inflected lexemes (“words”), semantic subdomains, three traditional grammatical annotations of the GNT, clauses, and the two types of clause complexes.¹⁰

Statistical Test and Design: Three-way contingency tables (language measure by authorship theory by genre theory) were analyzed by Hierarchical Log Linear Analysis. The experimental design included five separate suites of HLLA analyses, totaling 7,128 models. In the first suite, 1,080 HLLA models were built (9 tested language measures x 5 bin sizes x 6 authorship theories x 4 genre theories). The five bin sizes were as follows: unbinned, 25B, 50B, 100B, and 200B (where ‘B’ refers to bins in which forms below the given frequency were combined into a final category). In the second suite, 2,160 HLLA models were built by *instance*

⁷ The idea here is to separate the effects so as to compare them without confounding them. Neumann to the best of our knowledge was the first to explore this quantitatively; Neumann, *The Authenticity of the Pauline Epistles*.

⁸ It will be recalled that a number of mixed genre corpora, when submitted to eigen-systems decompositional methods, have demonstrated that the major moment of variation is genre rather than authorship. See especially Burrows, “The Statistical Analysis of Narrative Style,” 64; Elliott and Valenza, “A Touchstone for the Bard,” 201; Baayen et al., “Outside the Cave of Shadows: Using Syntactic Annotation,” 121–22. In Tweedie’s words: “In their study of syntactic data, Baayen/van Halteren/Tweedie (1996), like Binongo find that genre overrides authorial differences.... critical essays by an author are more similar to other critical essays than they are to crime fiction written by the same author.” Tweedie, “Statistical Models in Stylistics,” 389.

⁹ The power of the answer to these three question when they are asked of HLLA must be fully appreciated. Unlike whole table level measures such as Cramér’s V, G&K’s tau, or UC, recall that HLLA has decomposed component variance into its separate effects—it disentangles authorship stylistic variance (A|G) from genre stylistic variance (G|A).

¹⁰ Word groups were dropped because of a near unimodal distribution. Major semantic domains were dropped because of very low overall association with any authorship or genre hypotheses and because these major domains are subsumed within semantic subdomains.

deciles¹¹ (10 deciles x 9 language measures x 6 authorship theories x 4 genre theories). In the third suite, 2,160 HLLA models were built by *category* deciles.¹² In the fourth suite, 864 HLLA models were built by *instance* quartiles. The design of the fifth suite was identical to the fourth suite except it was run by *category* quartiles.

¹¹ *Instance deciles* contain the same number (or very close to the same number) of total *instances* in the GNT. There are for example 138,019 instances of words (inflected lexemes) in the GNT.

¹² *Category deciles* contain the same number (or very close to the same number) of total *categories*. There are for example 17,736 categories of words (inflected lexemes) in the GNT.

7.2 Hypothesis One: Is Authorship or Genre “Stronger” Overall?

Table 7.1 summarizes the results of all 216 unbinned HLLA models built (nine language measures by six authorship theories by four genre theories). These models constitute the *baseline set* since this set includes all language instances of all nine tested language measures in the GNT. (It is termed the baseline set because no low-frequency information is obscured by binning or eliminated by partitioning.) For each cell in Table 7.1, nine models were built, one for each language measure, with the number representing the total HLLA models where authorship (A|G) was “stronger” than genre (G|A) per the definition just given. Hence, the “4/9” at the intersection of Genre 8 and Modern Critical 18 means that for four of the nine tested language measures authorship (A|G) was stronger than genre (G|A).

	Genre 13	Genre 12	Genre 10	Genre 8	Row Totals
Baur Original (21)	0/9	0/9	0/9	4/9	4/36
Baur Reduced (19)	0/9	0/9	0/9	4/9	4/36
Modern Critical (18)	0/9	0/9	0/9	4/9	4/36
Modern Critical (17)	0/9	0/9	0/9	4/9	4/36
Modern Critical (15)	0/9	0/9	0/9	4/9	4/36
Traditional Authorship (9)	0/9	0/9	0/9	0/9	4/36
Column Totals	0/54	0/54	0/54	20/54	20/216

Table 7.1: HLLA Models in Which Authorship Accounted for a Larger Component of Variance¹³ than Genre: Unbinned Data

¹³ Per our definition of strength in the text, in all 29 models where authorship was stronger than genre, both terms were removed in the same step. All 216 models can be inspected at: ftp://decisionssupportsciences/stat_ling/HLLA_GNTRResults_RCS_Instance.xlsm.

Discussion of Hypothesis One Findings (Table 7.1):

Six primary findings are noteworthy:¹⁴

1. No authorship theory (A|G) (0 out of 54) was as strong as the 13-category genre theory.
∴ No authorship theory, ancient or modern, accounted for as much variance as the 13-category genre theory on any of the nine language categories tested. Simply stated Genre 13 swamps authorial variation. Translating these findings back to the older terminology of “markers,” either (a) many more markers (by category) reflect genre rather than authorship in the GNT or (b) a given marker is instantiated (chosen) more frequently for reasons of genre than authorship or (c) some combination of both. Ceteris paribus, unless prior studies have controlled for genre variation such as we have here, it is probabilistically likely that their “markers” were demonstrating genre differences rather than authorship differences.
2. No authorship theory (A|G) (0 out of 54) was as strong as the 12-category or 10-category genre theories. It needs to be underscored, though, that the reduced chi-squared value for G|A over A|G was not as high as was seen in the 13-category genre theory (point 1 above).
∴ This carries the identical conclusion as the point immediately above. Thus there are three genre theories that have “run the table” against all six authorship theories.
3. Overall, authorship (A|G) was stronger in only 9.25% (20 out of 216) of the models.
4. When the various genre theories sacrificed either type, audience or number (e.g. as they proceeding from Genre 13 to Genre 8) the strength of genre (G|A) over authorship (A|G) decreased. *This implies that type, and audience and number are critical constitutive characteristics of genre in the GNT.¹⁵*
∴ The implication above follows because as we proceed from Genre 13 to Genre 8 we are combining together books by either combining types, audience, or numbers. See Table 5.5. (An example of combining types is adding in John with the Synoptics as Genre 10 combines to become Genre 8.)
5. Models in which authorship (A|G) was stronger than genre (G|A), only occurred with the weakest genre theory (Genre 8).
∴ It is only when we “hamstring” genre that it weakens enough for authorship to begin to compete with it.

¹⁴ The complete data set can be inspected electronically in the spreadsheet at: ftp://decisionssupportsciences/stat_line/Design_1_HLLA_9Measures_by_6Authors_by_4Genres_by_No_Bins_Partial_Association.xlsx.

¹⁵ In simplifying from Genre 10 to 8 in particular, the consensus of mainstream critical thought would view Philemon as a letter to an individual and the Pastorals as pseudepigrapha written to the second century church. Therefore proceeding from Genre 10 to Genre 8 is combining not only types (John with the synoptics) but combining audiences *and* number (Philemon with the Pastorals).

7.3 Hypothesis Two: Is Authorship or Genre More Associated with *Infrequent Data*?

Both binning and partitioning designs were implemented to test the pseudepigraphal assumption that authorship is differentially expressed via low-frequency forms. This yielded 32 *other* tables of 216 HLLA runs with the same design as the baseline set presented in Table 7.1. Four of the tables had different bin designs, and 28 tables reflected different partitions of the data (10 frequency deciles, 4 frequency quartiles, 10 frequency deciles, and 4 category quartiles). To best summarize a dataset this large (7,128 total HLLA models) two representative exhibits were selected. In the first exhibit (Table 7.2) the lowest quartile instance partition, the 25% of the linguistic instances that occur least frequently is presented. In the second exhibit (Table 7.3) the strength of authorship and genre across the four quartile partitions are presented by quartile instance.¹⁶

	Genre 13	Genre 12	Genre 10	Genre 8	Row Totals
Baur Original (21)	0/9	0/9	0/9	1/9	1/36
Baur Reduced (19)	0/9	0/9	0/9	1/9	1/36
Modern Critical (18)	0/9	0/9	0/9	4/9	4/36
Modern Critical (17)	0/9	0/9	0/9	4/9	4/36
Modern Critical (15)	0/9	0/9	0/9	4/9	4/36
Traditional Authorship (9)	0/9	0/9	4/9	0/9	0/36
Column Totals	0/54	0/54	0/54	14/54	14/216

Table 7.2: HLLA Models in Which Authorship Accounted for a Larger Component of Variance than Genre: The Least Frequent Language Data (lowest 50% by frequency) in the GNT

¹⁶ The complete data set can be inspected electronically in the spreadsheet at: ftp://decisionssupportsciences/stat_ling/Design_5_HLLA_NineLinguisticMeasures_by_6Authors_by_4Genres_by_FreqInstanceBinnedQ4.xlsx.

Discussion of Hypothesis Two Findings (Table 7.2):

In comparing Table 7.2 (the lowest 25% by frequency) to Table 7.1 (all the data) all the five primary findings from Table 7.1 are also true of Table 7.2. In addition, however, *the genre theories are even stronger in the lower frequency partitions of the data*. Specifically, authorship (A|G) weakens in strength with low-frequency data, dropping from 20 models in Table 7.1 to 14 models in Table 7.2.¹⁷

To test this surprising association between low-frequency data and genre further we explored both low- and high-frequency binning and partitioning designs in the 31 remaining tables of 216 HLLA runs that had not yet been explored. In short this pattern continued consistently; A|G consistently weakened in the lower frequency bins and partitions of the data. In one particularly noteworthy example, when the data was examined with category quartiles, authorship (A|G) dropped to 6 and 0 models out of 216 for the second to lowest quartile and lowest quartile by category respectively.¹⁸

The experiments above collectively demonstrate that authorship (A|G) progressively *weakens* with lower frequency data partitions compared to genre (G|A). ∴ *This directly and thoroughly contradicts the oft-repeated dictum that low-frequency data is disproportionately authorial*. To test this finding by contrast, we next focused on the *high-frequency* partition of the data. As was to be expected, with higher frequency data partitions authorship (A|G) becomes stronger compared to genre (G|A). Below are the highlights of that work:

1. For the Top 50% of forms by frequency, authorship (A|G) was stronger than genre (G|A) in 70 out of 216 models (32% of the models).¹⁹

¹⁷ See the worksheet labeled "Authorship_vs_Genre_RCS_Instance" in the spreadsheet at: ftp://decisionssupportsciences/stat_ling/HLLA_GNTRResults_RCS_Instance.xlsm.

¹⁸ See the worksheet labeled "Authorship_vs_Genre_RCS_Category" in the spreadsheet at: ftp://decisionssupportsciences/stat_ling/HLLA_GNTRResults_RCS_Category.xlsm.

¹⁹ See the worksheet labeled "Authorship_vs_Genre_RCS_Instance" in the spreadsheet at: ftp://decisionssupportsciences/stat_ling/HLLA_GNTRResults_RCS_Instance.xlsm

2. When all authorship theories are compared to the strongest genre category, Genre 13, however, that number dropped to 11 of 54 (20.3%) of the models.²⁰
∴ Genre 13 is still “strong enough” to be stronger in almost 80% of the models even with high-frequency partitions of the data.

In summary, the progressive weakening of authorship for the least frequent forms directly contradicts the early twentieth-century notion (by Harrison and others) that authorship is differentially expressed by low-frequency forms such as *hapax legomena*. In the GNT at least, the converse is true; when disentangled from each other, *genre (G|A)*, *not authorship (A|G)* is differentially expressed “stronger” with low-frequency forms.²¹

In concluding this section and its associated hypothesis the question that looms large is this: *what might explain how the GNT became populated with so many low-frequency genre-associated forms?* The answer must draw from two fields: SFL theory to provide the contextual/linguistic insight necessary and corpus linguistics to provide the corpus/distributional insight. The explanation is as follows. It is granted on all sides that the New Testament authors/editors/redactors created the texts of the NT to address a broad and variegated set of needs within the theological communities they sought to serve. Said linguistically, in each occasion the author/editor/redactors were responding to a context of situation that called forth a given genre/register required for that context.²² Said more colloquially, specialized contexts call upon specialized registers—replete with their often quite rare lexical, semantic, and syntagmatic resources. This is all clear enough. We now need only one more piece to complete the puzzle,

²⁰ See the worksheet labeled “Authorship_vs_Genre_RCS_Instance” in the spreadsheet at: ftp://decisionssupportsciences/stat_ling/HLLA_GNTResults_RCS_Instance.xlsm.

²¹ This is not to be taken to mean that low-frequency idiolectic (especially idiosyncratic) linguistic species reflective of authorship do not exist in corpora! They assuredly do, especially in homotypic corpora. Nor are I saying such species do not exist in the GNT. The argument here is a statistical one—authorship-related low-frequency forms do exist in the GNT, but the data indicates that they are in the minority compared to those associated with genre.

²² Halliday, in one of his more colloquial moods, defines genre/register as “a kind of variation in language that goes with variation in the context of situation.” Halliday and Hasan, *Language, Context, and Text*, 38. See also Semino, “Stylistics and Linguistic Variation,” 41–44.

and it is this: there is an implication to the truism that the NT is a mixed genre, deeply-variegated occasional corpus. That is, when the number of occasions and contexts becomes quite large, such as that found in mixed-genre corpora, an interesting distributional shift occurs. Even *high-frequency* linguistic species (e.g. words, word groups or clauses) in a single register become *low-frequency* when that register is found rarely in the corpus. The register of baseball may serve as an example. Like all registers it is characterized by specialized resources—such as the collocations “squeeze bunt,” “steal sign,” and “double play.” The fan who gets to a ball game only every five years may use those collocations, say, a dozen times in that span—and likely all on the day that she or he attends the game! Similarly, an entire mixed corpus becomes populated by specialized linguistic species from dozens or hundreds of contexts—and these instantiated “species” which are found in high frequency in their specialized registers become low-frequency as that register competes with dozen or perhaps hundreds of other registers in the entire body of a mixed-genre corpus such as the GNT. As will be discussed later, this logic extends of course down to *hapax legomena* themselves.

7.4 Hypothesis Three: How Do the Strongest Theories Compare?

In hypothesis two we executed separate HLLA experiments using four exclusive quartile partitions by instance, and reported the result of the lowest frequency partitions.²³ In this design the focus was upon pairing the strongest genre theories against the strongest authorship theories. Determining the strongest genre theory was quite straightforward because Genre 13 consistently demonstrated higher HLLA performance relative to authorship. The strongest authorship theory, however, was far less clear given that authorship theory strength varied by frequency. Hence all authorship theories were included. This yielded an experimental design with 216 models (1 genre theory x 6 authorship theories x 9 language measures x 4 partitions.). The results were as follows:

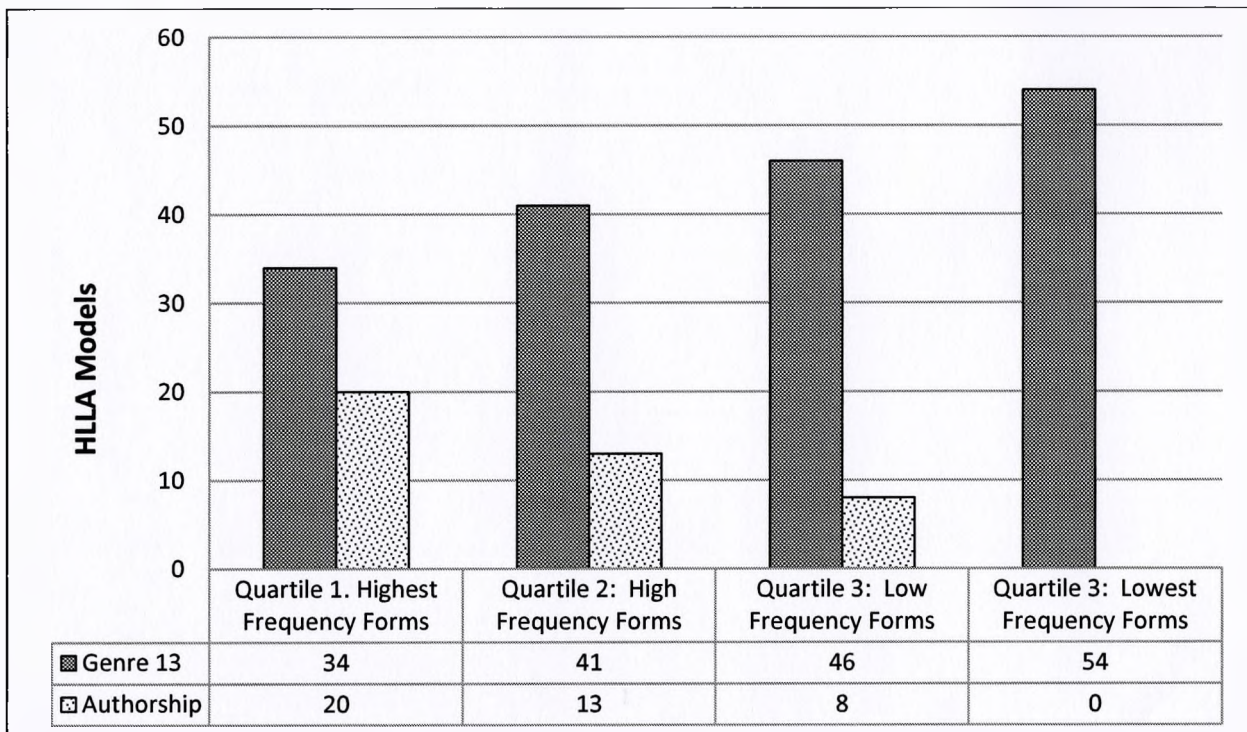


Table 7.3 Authorship vs. Genre as a Function of High- vs. Low-Frequency Forms in the GNT²⁴

²³ The top 25% of forms by frequency were assigned to Quartile 1, the next 25% to Quartile 2, and so forth.

²⁴ See the worksheet labeled "Authorship_vs_Genre_RCS_Instance" in the spreadsheet at:

ftp://decisionssupportsciences/stat_ling/HLLA_GNTResults_RCS_Instance.xlsm.

HLLA Models In Which Authorship (A|G) Explained More Variance Than Genre (G|A) Genre 13 Models Only

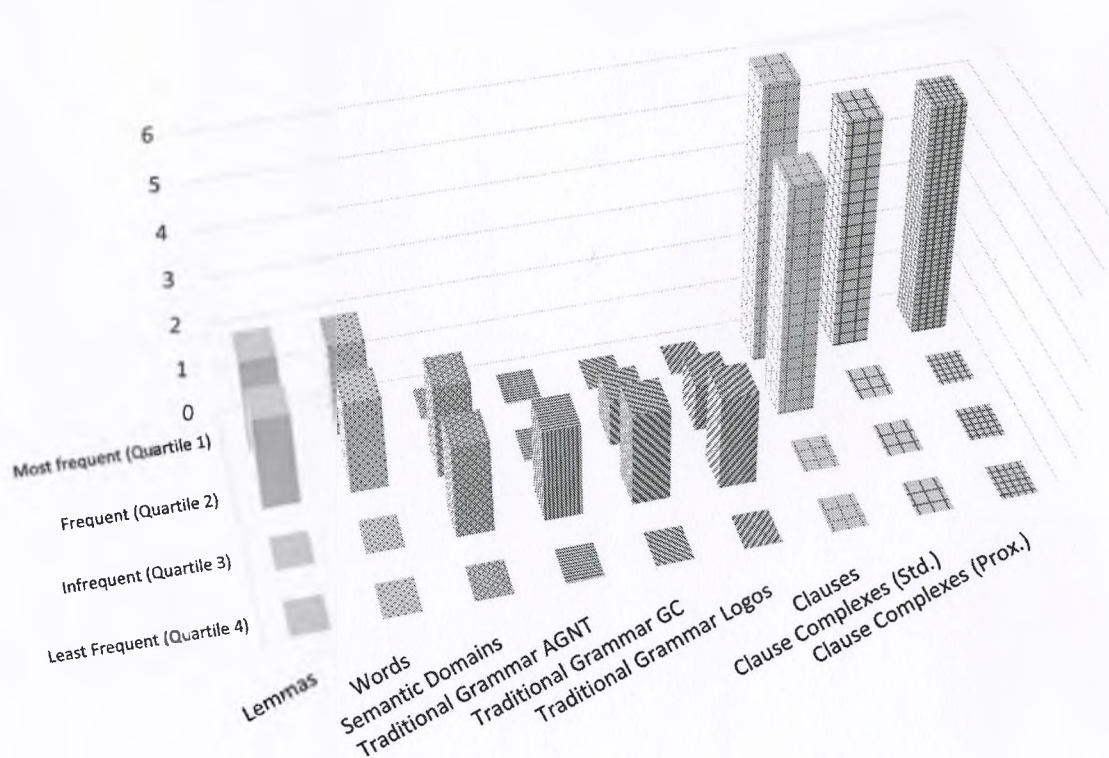


Table 7.4 Authorship Models That Explain More Variance Than Genre Models: by Partition and Language Measure

Discussion of the Findings from Hypothesis Three (Tables 7.3-7.4):

The partition data clearly demonstrates that Genre 13 became stronger with partitions that included lower-frequency language instances while the authorship theories (A|G) became weaker (Table 7.3). The general shape of the overall chart did not vary whether those partitions were by instance or by category²⁵ In short, as was seen in hypothesis two, low-frequency information is

²⁵ Genre 13 also explains more information when paired head-to-head with each of the authorship theories. Of the 36 HLLA models tested per authorship theory (9 linguistic measures by 4 quartiles), Genre 13 explained more variance in 21 of 36 models compared to the Original Baur hypothesis, 23 compared to the Reduced Baur hypothesis, 32 compared to Mainstream authorship (18), 32 compared to compared to Mainstream authorship (17), 32 compared to Mainstream authorship (15), and 35 compared to Traditional Authorship. This means that Genre 13 explains more

more associated with genre (G|A) than authorship (A|G).²⁶ Most intriguingly, however, when Table 7.3 is disaggregated by language measure (as in Table 7.4) more nuanced insight is gained. Specifically, Table 7.4 includes only those 41 HLLA models (out of 216) in which authorship (A|G) was stronger than genre (G|A). Thus Table 4.1 can be interpreted as the “places” researchers should explore to find markers for authorship. Several findings are noteworthy:

- 1) The traditional “locale” researchers have explored to transduce authorship—low-frequency words—is one of the poorest places to look for such markers.
- 2) One of the best places to discover authorship markers seems to be in high-frequency fractions of language at the rank of the clause and above.
- 3) Intriguingly the only places where traditional grammar (our proxy for paradigmatics) and semantics yielded authorship models that explained more variance than genre are in the middle two quartiles of the frequency distribution.

Overall these different “sweet spots” reveal that not only do different ends of the rank scale function differently in terms of the information they deliver over to us idiolectally vs. sociolectally, but that these linguistic species appear to operate differently in different fractions of the frequency distribution.

Because these insights cut cross-grain to the history of NT interpretation (specifically the assumption that authorship is encoded by low-frequency forms) I chose to cross-check the data as follows. The *least frequent* 1,000 clause proximity complexes in the GNT (out of 12,260) were chosen for analysis (all were clause complex *hapaxes*). The books in the GNT which had

variance in the GNT than any theory of authorship, but it also means that in 41 out of 216 models *when authorship is stronger*, the rank ordering represents the authorship theory that is most dissimilar from Genre 13. For the necessary parallel between this profile of results and the UC coefficient, see Krippendorff, *Information Theory*, 89.

²⁶ This was cross-checked as follows. The least frequent 1,000 clause proximity complexes in the GNT were chosen and cross tabulated against the books of the GNT. Nine texts of the GNT displayed a frequency of these clauses 1.2 times greater than the average Romans, 1 Corinthians, 2 Corinthians, 2 Thessalonians, Philemon, James, 1 Peter, 2 Peter and Jude.

the *highest frequency* of these *hapaxes* (a la Harrison’s practice)²⁷ were Romans, 1 Corinthians, 2 Corinthians, 2 Thessalonians, James, 1 Peter, and 2 Peter. These of course, all derive from one genre: *epistles*, and more specifically from a subgenre, *corporate epistolary*.²⁸ Further inspection reveals that these *clause complex hapaxes* were infrequent for a singular reason: their clause chains were very long. From a discourse perspective, long chains of clauses typically achieve paraenetic²⁹ or persuasive goals in discourse—which of course, is one of the signal functions of corporate epistolary genre in the GNT. Moreover, by implication *any* very long clause complex made it probable that the preceding clause was also an *hapax*. (Recall that clause complexes are defined from a head term—they start from the same first clause and subsequently branch from that primary clause.) *For this reason, these long chain, clause complex hapaxes*³⁰ clustered next to one another in eleven passages in the GNT. I chose to inspect nine of the longest of these clause complexes to discover if they shared any similarities. The nine clause complexes and an initial set of discourse similarities can be found in Table 7.5.

***Long Length Syntagms Identified by HLLA Analysis:
An Analysis by Discourse Function and Discourse Devices Used***

Text	Depth	Clause Complex (Proximity Encoded)	Discourse Function	Discourse Devices Used ³¹
Romans 5:3-5	8	1conjconjPAA-2P-3conjSCP-4SconjC-5SconjC-6SconjAP-7conjSPAA-8PC:2-2A:8-2A	Paraenesis: through the use of a catalog of clausal subjects (i.e. suffering: endurance-> character-> hope)	A sequence of five <i>secondary clauses</i> . Discourse devices: repetition (a core cohesive device) of each primary subject twice across clauses
Romans 6:9-10	9	1conjconjPAA-2P-3conjSAAP-4PA-5SCAP-	Paraenesis: to consider one’s self	A sequence of five <i>secondary clauses</i> . Discourse devices: five

²⁷ These had a frequency 1.2 times greater than the average frequency in the GNT.

²⁸ The *individual* epistles, specifically the Pastoral Epistles displayed significantly *fewer* clausal *hapaxes*, at a rate of 0.451, 0.364, and 0.488 the average for 1 Timothy, 2 Timothy and Titus.

²⁹ Unlike Stowers, I use paraenesis here to embrace both its protreptic (to encourage toward a new way of life) and its paraenetic senses (to encourage toward continuing the current way of life). This is closest to its ancient usage. Cf. Stowers, *Letter Writing in Greco-Roman Antiquity*, 92.

³⁰ Recall that this exercise was run only upon the bottom 1,000 *hapaxes*. I expect another 20 to 30 similar passages would emerge from this analysis if they were run on the remaining *hapaxes*.

³¹ For a brief summary of cohesive devices in NT discourse analysis, see Reed, “The Cohesiveness of Discourse,” 32–45.

		6CconjCPA-7CP-8CconjPC-9CP:3-2A:5-1A:8-1C:9-1C	dead to sin but alive to God. This is a discourse peak in Romans 6.	semantic chains that include a coextensional componential tie using repeated antynomy between θάνατος and ζωή
Romans 7:2-3	10	1conjSPCA-2AP-3SconjCPA-4P-5PA-6CP-7CPAA-8PSCA-9PA-10conjconjPS:3-1A:5-1C:9-2A:10-1A	Persuasion: freedom from the law using, by analogy, the end of a marriage by the death of a spouse	A sequence of fourteen <i>secondary clauses</i> . Discourse devices: two elaborative organic ties, repeated use of antynomy between θάνατος and ζωή. Lexical repetition (άνήρ [7x])
2 Cor. 7:7-11	16	1conjconjA-2APAA-3PCCCC-4PSA-5AP-6AP-7AconjP-8conjconjPA-9PconjA-10SconjCAP-11SconjCP-12PconjC-13SCPC-14PA-15APC-16SCPA:	Persuasion: regarding the sorrowful letter	A sequence of seventeen <i>secondary clauses</i> . Discourse devices: four causal-conditional organic ties, lexical repetition with (Λυπέω [8x], Μετάνοια [3x])
James 2:15-16	7	1conjSCP-2conjC-3PC-4PconjSC-5PA-6P-7conjP:4-1C	Paraenesis: by storytelling; a narrative of the hypocrisy of ignoring the poor	A sequence of seven <i>secondary clauses</i> . Discourse devices: repeated use of the imperative and subjunctive (six out of nine verb-forms)
1 Peter 2:21-24	9	1conjconjSPAA-2CPC-3conjPC-4SAAP-5P-6AAP-7P-8PconjC-9PA:3-2A:6-1A:8-1A:9-1C	Persuasion: through a catalog of Christ's submission to authority unto death—a discourse peak of the central Petrine theme of suffering	A sequence of ten <i>secondary clauses</i> . Discourse devices: three causal-conditional organic ties, three co-referential ties referring to Christ, four examples of abnegation (via the polarity system)
2 Peter 2:4-8	15	1conjconjSCAP-2P-3conjAAPA-4AP-5conjCAP-6conjACPA-7CAP-8conjCAAPA-9AP-10PC-11P-12conjCP-13PA-14AconjSAAAAP-15PA:3-1C:5-3	Persuasion: an extended <i>a fortiori</i> argument (angels, ancient world, Sodom, Lot) to incite obedience	A sequence of seven <i>secondary clauses</i> . Discourse devices: three repetitions of conditional organic ties (εἰ/καί paired with ἀλλά).

Table 7.5

Inspection of Table 7.5 yields several insights. *First*, it confirms the hypothesis of this hand-checked exercise. The low-frequency fraction of clause complexes are indeed disproportionately found in genre rather than in texts grouped by any known theory of authorship. *Second*, it confirms that *functionally similar* sections of NT discourse sections can be yielded *starting* from computational stylistics. What we discover in this regard is quite intuitive. This is appraisal language which gives rise to structured arguments which, in turn, use as one of their central devices long syntagms. In the NT this translates to long chains of clauses with a

deep variety of patterning and branching involving secondary and embedded clauses. *Third*, it demonstrates that computational stylistics, precisely because it “sees” things we do not see, may be used as a resource in NT discourse analysis. Although computational stylistic identification of discourse insights in the GNT is quite novel, one wonders what insights might be developed using other language measures (e.g. lexical, paradigmatic, syntagmatic and semantic), other language partitions (by frequency, variance and the like) and multiple combinations of these searched simultaneously. More can be added to this, but it draws us too far afield from our central point. I have demonstrated that *the high prevalence of low-frequency clause complex hapaxes are not found within authorship groupings as much as they are found in genre, and in one subtype of genre specifically, corporate epistolary.*

7.5 An Abductive Finding: An Authorship Pattern in the HLLA Deletion Statistics?

Finally, I present an intriguing pattern discovered in the course of inspecting the deletion parameters in the HLLA models. This was not an hypothesis of the study but was discovered in the process of the analysis. To demonstrate this pattern I need only present the deletion parameters for the first 30 of 216 of the unbinned HLLA (Genre 13) models.³²

Language Measure Tested	Categories (Bin Size)	Authorship Theory Tested	Genre Theory Tested	Effect Deleted First	Deletion Parameters		
					Chi-Sq.	df	CS / df
Lemmas	All (5413)	Original Baur (21 authors)	Genre (13)	Author	18046.40	108240	.167
Lemmas	All (5413)	Reduced Baur (19 authors)	Genre (13)	Author	15103.60	97416	.155
Lemmas	All (5413)	Mainstream (18 authors)	Genre (13)	Author	16533.73	92004	.180
Lemmas	All (5413)	Mainstream (17 authors)	Genre (13)	Author	15683.49	86592	.181
Lemmas	All (5413)	Mainstream (15 authors)	Genre (13)	Author	15666.07	75768	.207
Lemmas	All (5413)	Traditional (9 authors)	Genre (13)	Author	12087.08	43296	.279
Lexemes (Words)	All (17736)	Original Baur (21 authors)	Genre (13)	Author	33945.90	354700	.096
Lexemes (Words)	All (17736)	Reduced Baur (19 authors)	Genre (13)	Author	28943.67	319230	.091
Lexemes (Words)	All (17736)	Mainstream (18 authors)	Genre (13)	Author	31384.19	301495	.104
Lexemes (Words)	All (17736)	Mainstream (17 authors)	Genre (13)	Author	29999.64	283760	.106
Lexemes (Words)	All (17736)	Mainstream (15 authors)	Genre (13)	Author	29974.79	248290	.121
Lexemes (Words)	All (17736)	Traditional (9 authors)	Genre (13)	Author	24313.20	141880	.171
Sem. Sub-Dom.	All (663)	Original Baur (21 authors)	Genre (13)	Author	7096.26	13240	.536
Sem. Sub-Dom.	All (663)	Reduced Baur (19 authors)	Genre (13)	Author	5801.71	11916	.487
Sem. Sub-Dom.	All (663)	Mainstream (18 authors)	Genre (13)	Author	6629.81	11254	.589
Sem. Sub-Dom.	All (663)	Mainstream (17 authors)	Genre (13)	Author	6231.66	10592	.588
Sem. Sub-Dom.	All (663)	Mainstream (15 authors)	Genre (13)	Author	6220.81	9268	.671
Sem. Sub-Dom.	All (663)	Traditional (9 authors)	Genre (13)	Author	4044.79	5296	.764
TGram. (AGNT)	All (663)	Original Baur (21 authors)	Genre (13)	Author	5609.14	13240	.424
TGram. (AGNT)	All (663)	Reduced Baur (19 authors)	Genre (13)	Author	4555.71	11916	.382
TGram. (AGNT)	All (663)	Mainstream (18 authors)	Genre (13)	Author	5423.73	11254	.482
TGram. (AGNT)	All (663)	Mainstream (17 authors)	Genre (13)	Author	5078.90	10592	.480
TGram. (AGNT)	All (663)	Mainstream (15 authors)	Genre (13)	Author	5069.61	9268	.547
TGram. (AGNT)	All (663)	Traditional (9 authors)	Genre (13)	Author	3422.40	5296	.646
TGram. (GC)	All (583)	Original Baur (21 authors)	Genre (13)	Author	5606.39	11640	.482
TGram. (GC)	All (583)	Reduced Baur (19 authors)	Genre (13)	Author	4552.19	10476	.435
TGram. (GC)	All (583)	Mainstream (18 authors)	Genre (13)	Author	5405.17	9894	.546
TGram. (GC)	All (583)	Mainstream (17 authors)	Genre (13)	Author	5065.49	9312	.544
TGram. (GC)	All (583)	Mainstream (15 authors)	Genre (13)	Author	5056.64	8148	.621
TGram. (GC)	All (583)	Traditional (9 authors)	Genre (13)	Author	3402.59	4656	.731

Table 7.6. The First 30 Genre (13) vs. Authorship HLLA models: Unbinned Data

³² Recall that the 216 unbinned models were described in Section 7.1 (9 measures by 6 authorship theories by 4 genre theories).

What Are Deletion Parameters and How Might They Be Used for Insights into Authorship?

Deletion parameters are parameters produced by stepwise multivariate models as they sequentially eliminate parameters (interaction effects or main effects) from their models. The parameter of most interest is the simple squared residual, the amount of variation *lost* when any given parameter is removed from the HLLA model. Recall that in conventional HLLA model building the central goal is *parsimony*: to adequately describe the model with the *least* number of parameters. Accordingly, the model sequentially eliminates parameters that account for the least amount of variance (or some other measure of fit) in its attempt to reduce the total parameter count. In our case, of course, our immediate goal is not parsimony but *relative explanatory power*. Again, using more colloquial language, even if the model throws them out, the NT computational stylist is deeply interested in picking two of them out and dusting them off—(A|G) and (G|A) in particular. If cross-model insights can be yielded regarding each of these terms, their relative magnitude may supply inferential insight into which authorship and which genre theories are the most primitive.

With this as background, the data in Table 7.6 can provide insights into authorship conclusions via the following five-fold rationale. First, as we have seen, the (A|G) term in an HLLA model controls for genre—it disentangles authorship from genre—and hence the various authorship theories can be compared to one another without the confounding effect of genre.³³ Second, since the deletion table includes the chi-square of *all* the removed effects, the normed

³³ I am fully aware that the deletion statistics being compared here are from different models—just as the findings in section 6.3 are developed from different tables. Three factors make cross-model HLLA comparisons—though not traditional—at least feasible in terms of yielding acceptable insights. First, the single parameter being deleted before the two way interaction terms of interest (A|G and G|A) is *always the same*, the three way interaction term. Second, the amount of variance in that three way interactions term is always very small. Third, two of the three dimensions of the model are always the same: the language measure and the genre theory. Fourth, as with the rationale for the cross-table comparisons in section 6.3, the chi-square between the authorship theory and its interaction with the language measure can be normed and therefore compared.

relative amount of variation (or normed variance) for the interaction between authorship and the language of the GNT (A|G) can be calculated as well as its associated statistical significance. Third, norming by the chi-square of the authorship (A|G) term, whether by phi, reduced chi-square, or Cramér's V, yields, by definition, a measure of the strength of association between that authorship theory and the language of the GNT. Fourth, it is reasonable, if not axiomatic, to posit that since the two-way interaction of authorship with a given language measure (A|G) removes the confounding effect of genre, the authorship theory that consistently displays the highest association with the language of the GNT is the one in which the authors are the most discriminated from one another. *Fifth, it is reasonable and even axiomatic in computational stylistics to posit that the authorship theory that consistently displays the most discrimination between its authors is, ceteris paribus, the best candidate for being the primordial theory of authorship.*³⁴ Given this rationale, Table 7.6 yields the following findings:

- When disentangled from the strongest genre categories (either the 13-category or 12-category genre theory) *the authorship theory that demonstrates the maximal normed variance is always traditional authorship (108 models out of 108 models).* This observation derives from the rightmost column in Table 7.6 which displays the chi-square statistic divided by the degrees of freedom (reduced chi-square). This observation holds true, it should be noted, whether the statistic employed is reduced chi-square, phi or Cramér's V.³⁵ [Note: The two most theoretically-grounded theories of genre, those that best respect the type, audience, and

³⁴ Colloquially stated, once we have subtracted out the effect of genre, what is "left over," all things being equal, is true authorial variation.

³⁵ See the highlighted values in the reduced chi-square (RCS), phi and Cramér's V columns in the Summary worksheet of the spreadsheet at:
ftp://decisionssupportsciences/stat_ling/HLLA_9Ling_by_6Author_by_4Genre_by_5bin_Summary.xlsx.

number differentiation we suggest in Section 7.1 are the 13-category and 12-category versions of genre. These will henceforth be collectively referred to as Genre†.]

- The data also confirms a reproducible hierarchy in the strength of the relationship between the various authorship theories. After norming by their respective degrees of freedom (reduced chi-square) inspection of the rank order of the authorship theories (A|G) in Table 7.6 reveals almost invariably the following rank order: (1) the traditional 9-author theory, (2) the reduced mainstream 15-author theory, (3) the reduced mainstream 17-author theory, (4) the original mainstream 18-author theory, (5) the original Baur 21-author theory, and (6) the reduced Baur 19-author theory, respectively.³⁶
- These initial results indicate that when authorship is allowed to “stand on its own” (controlled for nonretrograde³⁷ genre effects), and when all instances of the nine tested language measures of the GNT are considered (none are binned together), traditional authorship explains more of the variance structure of the language of the GNT than any other authorship theory. This was also true of 215 of 216 HLLA models built on frequency partitions of the data.³⁸
- When the findings were analyzed by bin, the hierarchical dominance of traditional authorship over the other authorship theories *decreases*. That is, traditional authorship increases its dominance over other authorship theories when infrequent forms are allowed to contribute to the model rather than be obscured by being recombined into a final category.³⁹ That, is, when

³⁶ See the Relative Strength of Authorship columns in the Summary worksheet of the spreadsheet at:

ftp://decisionsupportsciences/stat_ling/HLLA_9Ling_by_6Author_by_4Genre_by_5bin_Summary.xlsx.

³⁷ It is vacuous to use retrograde (the worst-performing) genre categorizations such as the Genre 10 and Genre 8 categorizations. Accordingly we have used the Genre 13 and Genre 12 categories in this study.

³⁸ The four partitions tested were quartiles 1-4 by frequency instances. In each partition 54 HLLA models were executed: 1 genre (13) x 9 linguistic measures x 6 authorship theories). See the Relative Strength of Authorship columns in the worksheets 1_25, 25_50, 50_75, and 75_100 within the spreadsheet at:

ftp://decisionsupportsciences/stat_ling/HLLA_9Ling_by_6Author_by_4Genre_by_5bin_Summary.xlsx

³⁹ As measured by both the likelihood ratio and Pearson chi-square.

infrequent forms are allowed to contribute to the models this strengthens the case for traditional authorship rather than weakening it. By inference, this means that once the traditional authorship theory (or any other authorship theory for that matter) has been disentangled from genre (A|G), *then and only then*, can the long-standing intuitive notion of the earlier stylometric researchers be countenanced—that low-frequency forms *do* encode idiolect choices—once the predominant low-frequency *genre* forms are removed.

- Another insight can be gleaned from Table 7.6. Because in all cases across all bins of data Genre 13 was the strongest genre theory, and because the other genre theories simplify the type/audience/number typology found in Genre 13, this infers that each dimension of the typology is important for genre. Because Genre 13, moreover, is the most detailed genre theory with regard to type, number and typology, its success begs the question regarding whether even more detailed genre categorizations (such as Genre 14 or 15) may be constructed and be stronger still. Such categorizations may be theoretically possible by creating a genre theory with more type categories⁴⁰ and/or assigning texts into more appropriate categories.⁴¹

What is the Overall Implication of this Pattern?

Per point four above, given that (a) the authorship theory that consistently displays the highest association with the language of the GNT (A|G) has the highest likelihood of being the primordial theory of authorship, and that (b) the authorship theory that meets this criteria is, without exception, the traditional theory of authorship (108 out of 108 times), then, by inference the traditional authorship theory is the best candidate among those tested here for being the

⁴⁰ Number and audience categories as we have defined them here cannot become more detailed.

⁴¹ Note: subsequent to this chapter, an inspection of correspondence analysis runs revealed a genre category that maximally separates genres via correspondence analysis. See section 8.5.

primordial theory of authorship. I am fully aware of how disruptive this finding may be to various mainstream critical theories of authorship. These and other issues will be addressed in the concluding section of this chapter.

7.6 Disentangling Authorship from Genre Using HLLA: Conclusions

Interpretively, the HLLA data permit eight conclusions.

First, once authorship (A|G) is isolated from genre (G|A), and the magnitude of their separate individual interactions with the language of the GNT is assayed, genre is demonstrably “stronger”⁴² than authorship (Table 7.1). Even more notably, when the strongest genre theories are compared to all authorship theories, no authorship theories were as strong as the strongest two genre theories (0 out of 108 HLLA models).

Second, whether the language of the GNT is explored via partitions, bins, or remains unbinned, low-frequency information is differentially associated with genre (G|A) not authorship (A|G) (Table 7.2).

Third, language measures appear to function somewhat differently, although in general low-frequency information is still differentially associated with genre (Tables 7.3 and 7.4).

Fourth, as measured by deletion statistics from HLLA models, once the various authorship theories (A|G) are isolated from genre (G|A), and the magnitude of their separate individual interactions with the language of the GNT is assayed, the strongest authorship theory is the traditional theory of authorship (Table 7.6).

Fifth, these findings may allow an answer to be proposed to the question asked in Section 6.4.3: What is the cause⁴³ of the differential destruction of information seen in the traditional nine-author grouping in the GNT? Recall that pseudepigraphal theories assume that the preponderance of measured unique information is due to authorial information. With the advent of computational stylistics and nonparametric multivariate models that can “subtract” genre from

⁴² Again per the definition of strength defined in section 7.1.

⁴³ We define “cause” here in its informal semantic sense, as formal causal models (LISREL, etc.) have not yet been employed against this effort. These will be implemented in a subsequent cycle through the abductive design.

authorship (A|G) this assumption can now be tested. The pseudepigraphal assumption necessitates that as increasingly low-frequency partitions of linguistic data are inspected, the interaction term between (A|G) authorship and language of the GNT should increase. *In fact authorship (A|G) decreases, while genre (G|A) increases (Table 7.1).* Within the constraints of our overall experimental design, this is our third independent line of evidence that unique (low-frequency) information in the GNT is more associated with genre than authorship. More on the implications of this finding will be covered in the seventh point below.

Sixth, these findings may allow us to propose an answer to the question posed in Section 6.4.3. Recall first that the uncertainty coefficient (UC) can be viewed as the proportion of unambiguous information (information received without either noise or equivocation) sent from a sender to a receiver. Recall as well that the traditional authorship theory displayed a far lower UC relative to the other theories of authorship. Finally, recall from this chapter that the traditional authorship theory (once it is disentangled from genre) is the *strongest* of the authorship theories. What, then, is the cause for the lower UC for the traditional authorship theory in Table 6.7? Especially in light of its higher (A|G) here? The answer relates, first, to what UC is actually measuring. The UC measures the proportion of *total information* in the grouping of the texts (since that is the only difference between the various theories) being unambiguously “communicated” in the contingency table. Thus something about *how* the traditional authorship theory is being grouped is differentially causing *information loss*, a lower proportion of total information communicated via the traditional nine-author theory. Information theory (chapter nine) defines that only two things can cause information loss: noise or equivocation. So which is it? Here is where the ability of multivariate analysis to disentangle main effects (such as authorship and genre) can lend us a hand. The only relevant difference between the differentially

higher (A|G) displayed by the traditional authorship theory in this chapter and the differentially lower UC seen in Table 6.7 for the traditional authorship theory is that, as we have mentioned quite frequently, HLLA mathematically removes genre to yield (A|G). Thus the only explanation for the differential loss of information is that the genre fraction still present in the UC value must—somehow—differentially cause either noise or equivocation. While work as to whether this loss is due to noise or equivocation is still to be fully executed, I argue that, theoretically, only equivocation can be the answer. (See the information theory primer in chapter nine.) Even if we should eschew mathematical theory to make this same point, there is a more accessible way to think about it. According to the traditional nine-author theory, one single author in the NT, Paul, wrote across *six* of the thirteen genres in the GNT, Luke wrote across *two* genres, John across *three* genres, and Peter across *two* genres. This combining of genres is far greater than in any other authorship theory. Because of the substantial interaction between authorship and genre, this is simply a recipe for equivocation—and on a massive scale. Said only slightly more mathematically, equivocation destroys information, and the theory that has the most equivocation between main effects with strong interaction *will* display the sharpest reduction in UC. Yet the story isn't over quite yet. Genre, per the data in this chapter, is the largest source of variation in the GNT and, moreover, it differentially expresses itself in the lowest frequency partitions of the language. This is exactly the partition to which information theory, by design, is most sensitive. Thus, the reason why information is being differentially destroyed in the nine-author theory, we submit, is because genres (not authors per the pseudepigraphal hypotheses) are being grouped together.

Seventh, what conclusions can be drawn regarding the various pseudepigraphal theories as viewed from the stance of the computational stylistic work executed in this chapter? What

would we expect to see if the pseudepigraphal theories best accounted for this data? Two answers are clear from the data in this chapter: (1) We would expect to see authorship (per P.N. Harrison) differentially encoded in the lower-frequency partitions of language. It is, in fact, differentially encoded in the *higher-frequency* partitions of language. The data, then, is precisely opposite of what pseudepigraphal theory predicts. (2) We would expect to see the differential destruction of *authorship* information. Instead we see the differential destruction of *genre* information. To understand this second point, recall that that pseudepigraphal theories of style view the traditional nine-author theory, in essence, as a mish-mash of pseudepigraphal authors. Depending on which mainstream critical view one holds, there are at least four and as many as ten hands that combined to write the thirteen texts of the Pauline canon (See Table 5.3). According to those who posit pseudepigraphic authorship then, the traditional view of a singular Paul *requires* that there is differential destruction/obscuring of *authorial style* rather than *generic style*. Given that authorship information (per Burrow's work and this study) is differentially encoded in the highest frequency fractions of language, one would then not expect much information destruction at all because information theoretic statistics (e.g. the Uncertainty Coefficient) are differentially sensitive to the destruction of low-frequency rather than high-frequency information. *In contrast, the opposite is observed.* The nine-author (traditional author theory) demonstrates the highest amount of low-frequency information destruction—exactly where the signal for genre is highest. Again, this is opposite of what one would expect if the pseudepigraphal theories were the best explanation of the data.

Eighth, parallel to the logic above, given that the strongest theory of genre is Genre 13 and given that, by the same logic, the genre theory (G|A) that explains the most variation in the text has the highest likelihood of being the primordial theory of genre, Genre 13 enjoys that

distinction. As before, however, it is too early in our experimental design to make an unqualified claim that this finding is certain.

Given these conclusions then, what historical setting best fits the data and theory just described? First, as measured by the deletion statistics, the most primordial authorship theory is the theory that displays the strongest interauthorial differences after genre effects have been removed. Second, given the “information destruction” argument in point five, the question reduces to which historical setting describes a relatively small number of redactors/editors/authors who write across multiple genres. Out of our set of six authorship theories the only theory that meets both these criteria is the traditional theory of authorship. The surprising nature of this finding, at least from the perspective of mainstream NT scholarship, is granted. Whether surprising or not, the question reduces to whether this finding has been developed adequately and whether further research, hopefully undertaken under a defensible experimental design, will further confirm it.

Four issues, in my view, should be clarified to place this discussion in context. *First*, we need to interrogate the method. Maximizing the robustness of multivariate findings requires that we understand the underlying variance and covariance structure of our data. In essence, we need to turn the data around and look at it from many angles. Indicated strategies for doing so include (a) binning the relevant categories of the contingency table and (b) exploring the model across a wide range of degrees of freedom by either eliminating or recombining categories by frequency fraction or by (c) exploring reproducibility by various schemes of validation. In terms of these requirements the design just executed has achieved tasks (a) and (b) and, in effect, by executing a principled and relatively large suite of bin and partitioning strategies, this study has achieved (c). *Second*, further methods of nonparametric multivariate analysis are indicated including, without

restriction, other forms of log-linear models, reparameterized models, restricted canonical and restricted association models.⁴⁴ *Third*, the issue of validity is not merely a statistical one, nor even a task for post-positivist validation, but rather it is also a quintessentially *hermeneutical* one. More on this will be discussed in Chapter Ten. *Fourth*, in abductive terms, even though this study has executed over 7,000 models, that still makes it relatively early on in a comprehensive experimental design. There are likely other interactions that were not measured and there are certainly other main effects that should be modeled. Candidates for those main effects include register variation, audience effects, style shift/drift, stylochronometric effects and the like. Ideally it would be preferable to develop those main effects *a posteriori* through an unsupervised EMVA exercise. I trust that even these brief comments underscore why abduction is so very critical.

With all the requisite “caveats” now covered, IBE still requires us to offer—with what is known now—the best explanation for the data. An analogy may suffice. If the contested issue of New Testament authorship can be compared to a horse race with our sequential experimental design as the racetrack, we are indeed in the early stages of our race⁴⁵—but the most ancient horse just took the lead.

⁴⁴ Multivariate models useful for the analysis of multiway contingency tables such as those described by Clogg (1982), Goodman (1986) Gilula and Haberman (1988) and Becker (1989) are suggested as a starting point. See Clogg, “Using Association Models in Sociological Research,” 114–34; Goodman, “Some Useful Extensions of the Usual Correspondence Analysis Approach,” 243–70; Gilula and Shelby, “The Analysis of Multivariate Contingency Tables,” 760–71; Becker, “Models for the Analysis of Association in Multivariate Contingency Tables,” 1014–19.

⁴⁵ Other structural modeling (e.g. IT-based structural modeling) approaches have not been run to confirm the initial findings here, nor have latent class models been run to inspect the possibility of unobserved variables within the variance/covariance structure of the GNT.

8 Results Part III: Visualizing NT Text Clustering Using Correspondence Analysis

8.1 Visualizing How the NT Texts Cluster:¹ An EMVA Primer

Chapters Six and Seven demonstrated that whether measured by association (Tables 6.1-6.6)² or by nonparametric multivariate (HLLA) modeling (Tables 7.1-7.4), genre is “stronger”³ than authorship across all nine classes of linguistic measures assayed so far in the NT. Those chapters, moreover, demonstrated two different aspects of the relationship between authorship and genre in the NT. The association-based statistics measured in Chapter Six, on the one hand, measured the total or *undecomposed* relationships in the data. All main effects and interaction effects from all sources (genre, authorship, occasion, theological setting, register, audience, etc.) were still “inside” each measure. The HLLA models in Chapter Seven on the other hand, separated or “disentangled” the effects of authorship (A|G) from (G|A). Given that genre is modestly stronger than authorship whether measured *in toto* or compositionally, this itself raises two questions:

1. How is this modestly higher strength of genre over authorship reflected in the relationship *between* the NT texts. That is, how do the texts of the NT “cluster”? If we could visualize these clusters, do they cluster by genre, authorship or something in-between? Does this clustering differ by language measure?
2. What can this grouping or clustering of the texts of the NT tell us about the relative strength of authorship and genre compared to other effects?

¹ Following corpus linguistic convention, henceforth we will refer to the NT books as *texts*.

² With the caveat that the uncertainty coefficient is far more accurately understood as a measure of normed mutual information (TI:(SR)).

³ Recall the technical definition of the term “stronger”: With bivariate association and bivariate IT-based measures, “stronger” means that the statistics for genre are higher in magnitude than for authorship. For HLLA modeling, it means that a) an interaction effect (e.g. G|A) remains in the HLLA model when another interaction effect (e.g. A|G) is removed, or b) the normed chi-square of one interaction effect is higher than the other when both are still in the model.

Fortunately, as we have already seen, the broad field of Extractive Multivariate Analysis (EMVA) exists to answer these kinds of questions. When EMVA techniques are used to visualize data they are often termed Multivariate Data Reduction (MDR) techniques. It will be recalled from Chapter Two that these approaches in their many forms simplify complex data relationships using eigen-system mathematics.⁴ In the world of MDR, this is exactly what is desired, because it allows complex datasets with many variables to be plotted onto a two- or three-dimensional “map” such that the maximal amount of variance is compressed into these “mappable” dimensions. When the type of data being analyzed is two-way or N-way contingency tables, the indicated MDR techniques are generally Correspondence Analysis (CA) and Multiple Correspondence Analysis (MCA), respectively.⁵ Both techniques are interpreted in the same way; categories that exhibit multivariate similarity are plotted closer to one another, and dissimilar categories are plotted farther away from one another. Such a map can place each row or column of the contingency table (e.g. the 27 NT texts and, for instance, the 50 most frequent lemmas of the GNT) as a point on that map.⁶

Using CA or MCA provides five distinct advantages in understanding the entangled relationship between authorship and genre:

- *Interpretability Through Visualization:* In the resulting CA or MCA maps, NT texts that are closer to one another are more similar to one another in relation to the language measure used to generate the map, and texts farther away are more dissimilar. This

⁴ Eigen-systems are widely used in the analysis of the large data sets found in data mining, psychology, sociology and marketing. For accessible introductions to eigen-systems in linear algebra see Lipschutz et al., *Linear Algebra*, 107–23; Golan, *The Linear Algebra a Beginning Graduate Student Ought to Know*, 215–48.

⁵ For the inaugural work on CA, and its adoption in the French-speaking world see, respectively, Hirschfeld, “A Connection between Correlation,” *passim*, and Benzecri, “Elaboration statistique de données sur ordinateur,” *passim*. For the classic monograph on CA see Greenacre, *Theory and Applications of Correspondence Analysis*, *passim*. For two very accessible introductions to its application see the separate works by Clausen, and Meulman and Heiser; Clausen, *Applied Correspondence Analysis*, 2–26; Meulman and Heiser, “SPSS Categories 21.0,” 46–66. Excellent separate monographs on MCA include those by Greenacre, and LaRoux; Greenacre and Blasius, *Multiple Correspondence Analysis*, 4–29; Le Roux and Rouanet, *Multiple Correspondence Analysis*, 1–67.

⁶ Points are actually the weighted geometric center (centroid) of all the objects that underlie it. Thus, the centroid αὐτός is the weighted geometric center of its 5,597 instances in the NT.

distance-based metaphor⁷ is not only intuitive, it yields clusters of texts that can be readily interpreted as more “authorship-like” or “genre-like.”

- *Elimination of Theory Bias*: Because the contingency table being analyzed is simply a language measure tabulated against the 27 texts of the NT, no authorship or genre bias is explicitly expressed in the table. That is, the contingency table itself, and each cell in it, expresses the combined effects of authorship, genre, and any other context of culture or context of situation effects that influenced the linguistic choices of the author/editor/redactor of those texts. Because the calculated final positions of the texts are unbiased in this way, the correspondence maps express most clearly the most dominant component, regardless of whether it is authorship, genre, or any other effect.
- *Appropriateness for a-Posteriori Analysis*: Once the CA or MCA analyses are generated and interpreted, further CA runs can then be executed on two-way contingency tables where authorship effects (both main effects and interactions) or genre effects (both main and interactions) are removed by HLLA.
- *Relative Comprehensiveness*: Because both CA and MCA maps express a relatively high percent of the total information in three dimensions,⁸ these tools provide a comprehensive visual understanding of the relationship between the texts of the NT and the given language measure.
- *Coherence with Mixed Methods (MM) Research*: In MM approaches the central notion is for the data to be inspected using a hybrid approach that integrates qualitative analysis with quantitative analysis.⁹ CA or MCA produces qualitatively interpretable maps through quantitative means.

An Important Interpretive Point: Per the design in Chapter Four, the data in this chapter will be analyzed using the 27 texts of the GNT, rather than multiple samples of equal size from each book (per, for instance, Mealand).¹⁰ Both approaches are legitimate but require different interpretive frameworks. In the former case texts with more words will have more mass, just as they do in the corpus.¹¹ Visually, then, they will make a contribution to the reduced space proportional to the total number of words in that text. In the latter case (per Mealand’s work) the interpretation is different; it is as if every text was the same size in the NT. Each approach has its

⁷ A number of distance measures are available to use in calculating distances in an N-space (a space with N dimensions). Options include Euclidean, squared Euclidean, chi-square, Mahalanobis, Ward’s, etc. For simplicity we will be using a simple Euclidean distance as our comparator. For a review of distance-based measures of similarity see Aldenderfer and Blashfield, *Cluster Analysis*, 24–28.

⁸ The CA runs in Figures 8.1 and 8.3 express slightly more than 50% of the total information in the contingency table.

⁹ See especially Tashakkori and Teddlie, *Handbook of Mixed Methods*, 256–81.

¹⁰ See especially Mealand, “The Extent of the Pauline Corpus,” 69.

¹¹ For an understanding of mass in correspondence analysis, see Clausen, *Applied Correspondence Analysis*, 9–26.

advantages and disadvantages, and each must be interpreted in terms of its own framework. In abductive cycle three I plan on a full suite of normed CA and MCA runs, similar to Mealand's but upon coherent rather than mixed measures of linguistic rank.

8.2 The Experimental Design Used to Perform the Correspondence Analyses¹²

The initial experimental design cycle for Correspondence Analysis (CA) was as follows:

Language Measures: Ten of the original eleven language measures were selected for this exercise: five syntagmatics measures (lemmas, inflected lexemes or “words”, clauses, and the two types of clause complexes), two non-lexicogrammatical measures (major semantic domains and semantic subdomains) and the three paradigmatic measures (three traditional grammatical annotations of the GNT).

Experiments Executed: 180 initial correspondence analyses (CA) were executed.¹³

Specifically, each of the ten language measures above were varied by two methods of feature set selection, AAVASR and APASR.¹⁴ (Feature set selection¹⁵ finds maximally different categories within a given linguistic measure, and AAVASR and AAPASR are two separate algorithms used to do so. By category here we mean, for instance, a specific lemma within the set of all lemmas.) Three depths of AAVASR and APASR were used. That is, AAVASR and APASR were used as FSE methods to discover the top 26, 51, and 101 maximally different categories for each of the ten linguistic measures. These top categories in turn were drawn from a pool that required any single category to occur at least 10, 25 or 50 times in the GNT respectively. (Thus the total number of runs executed was 190: 10 measures by 2 feature set extraction methods [FSE's] by 3

¹² Portions of sections 8.2, 8.3 and 8.4 are substantially similar though not identical to a series of CA experiments that have submitted for publication in and will be forthcoming.

¹³ This section displays a subset of work currently being submitted for publication.

¹⁴ AAVASR = average absolute value of the adjusted standardized residual. The higher the AAVASR the more variant that linguistic measure is across the texts of the GNT. APASR = average positive adjusted standardized residual. APASR is calculated essentially the same as AAVASR except it uses only texts with positive residuals.

¹⁵ Feature set selection is sometimes termed feature set extraction.

top measure cutoffs by 3 minimal frequency sizes.) Note that the “pool” size submitted to the CA analysis was always one category greater than the frequency (e.g. 26 rather than 25) because a “Rest” category was constructed to contain all the forms that do not fit that criteria. This additional category yielded two explicit benefits: (1) the entire GNT was thereby included in each correspondence analysis executed and (2) all analyses at a given cutoff yielded a contingency (cross tabulation) table that had the same degrees of freedom for all ten linguistic measures tested.¹⁶ This allowed the visual display of the data to be more directly interpretable across linguistic measures.

Layout of Results: With a set of results this large (180 CA analyses) some method of briefly summarizing the results is clearly needful. Three exercises are suggested. First, an introductory CA run will be explored that will orient us to the interpretation of the remaining 180 CA runs in the design (Section 8.3). Second, three of the 180 runs will be selected to allow the CA data to be syntagmatically, paradigmatically and semantically compared and contrasted (Section 8.4). Third, the entire set of CA runs will be interrogated to help identify whether authorship or genre is “stronger” by inspecting which is differentially expressed in the lowest components of the CA extracted.¹⁷ (Section 8.5).

¹⁶ The degrees of freedom are as follows: 650 for the 26 by 27 table, 1,300 for the 51 by 27 table, and 2,600 for the 51 by 27 table.

¹⁷ Recall that in EMVA the first component explains the highest amount of variance, the second component the second highest, etc. Thus, if a given cause (sociolectic or idiolectic) is expressed in the first dimensions it explains more of the variation than if it is expressed in subsequent dimensions.

8.3 An Exercise in Interpreting Correspondence Analysis

Three orienting exercises will be executed in this subsection: the display and interpretation of the dimensions of the CA uniplots, a demonstration of the relationship and distances between large texts in the GNT, and a calculation of the best fit between the CA data and theoretically and empirically derived theories of authorship and genre.

Interpretive exercise #1 The display and interpretation of the dimensions of CA uniplots.

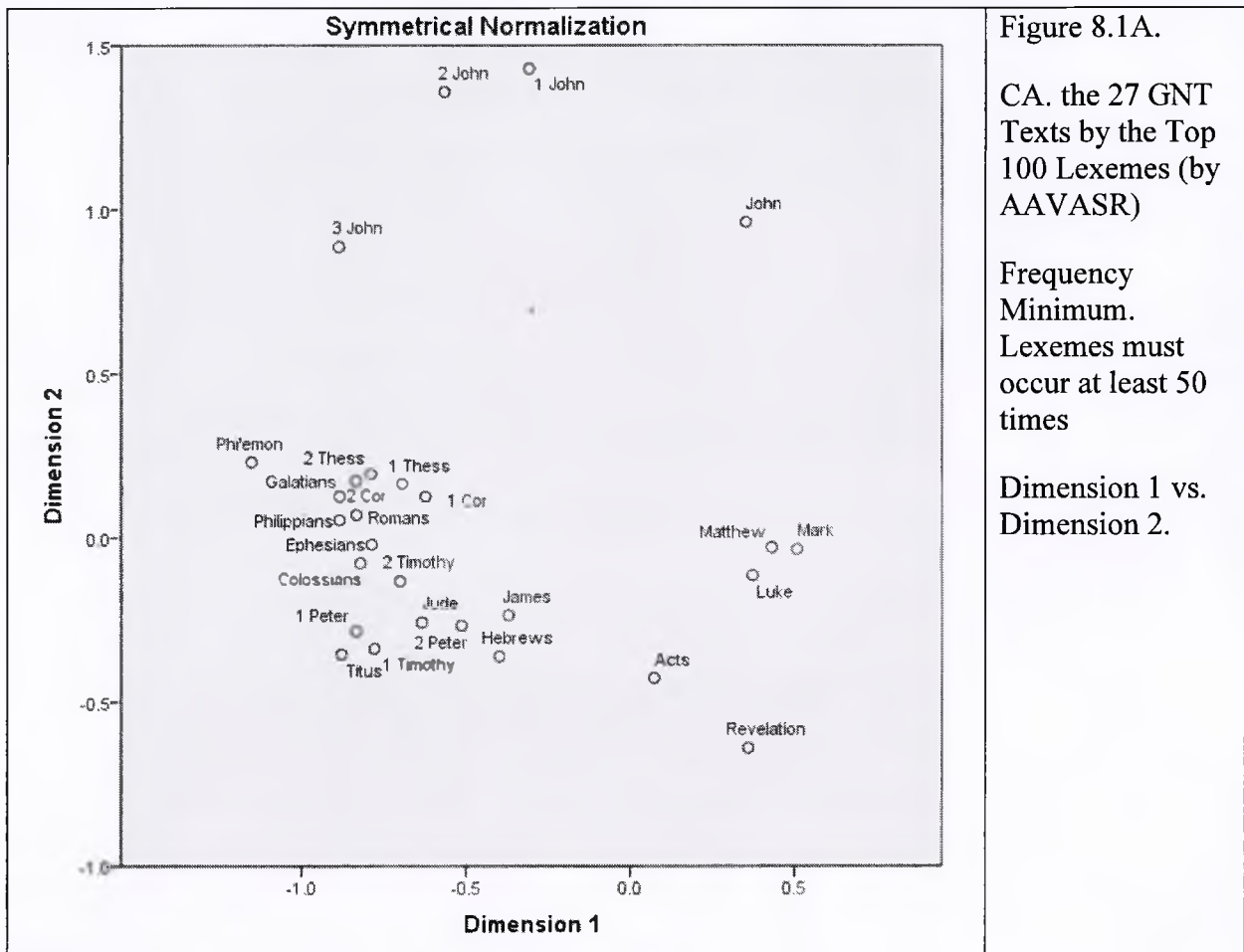


Figure 8.1A

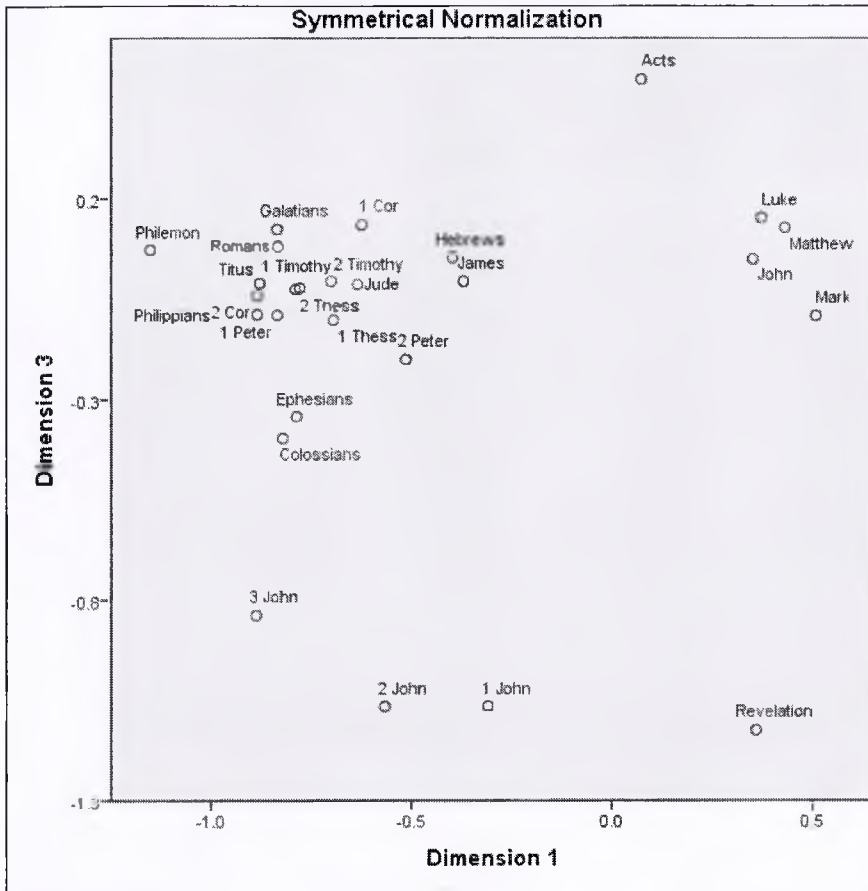


Figure 8.1B.

CA. the 27 GNT Texts by the Top 100 Lexemes (by AAVASR)

Frequency Minimum. Lexemes must occur at least 50 times

Dimension 1 vs. Dimension 3

Figure 8.1B

Inspection of both projections (Figures 8.1A and 8.1B) reveals that they both separate the NT into three complexes, and three texts that stand apart from the rest.¹⁸ The three complexes include a loose clustering comprised of 1-3 John (seen best in Figure 8.1B), a tight complex including the Synoptic gospels, and another somewhat diffuse complex of epistles. In addition, Revelation, Acts and the Gospel of John in particular, stand distinctly apart from these complexes and from one another (best seen in Figure 8.2 below). The difference between Figure

¹⁸ The CA's in this work were all initially created using IBM's SPSS software package. Rotation and other necessary analytics were performed by PositionSolve™, proprietary eigen-systems-based data reduction and visualization software originally written by the author and subsequently enhanced by Decision Support Sciences.

8.1A and Figure 8.1B is due simply to the fact that Figure 8.1A was rotated “upwards” around the horizontal axis 90 degrees thus fully revealing the third “depth” dimension.¹⁹

Interpretive exercise #2 Demonstration of the clustering and distances between the largest texts in the GNT



Figure 8.2. CA of the 27 GNT Texts by the Top 100 Lemmas by AAVASR. (The output is rotated so that texts lie as closely as possible upon the plane of the page.)

For this second interpretive exercise, inspect Figure 8.2 and recall that all but the most trivial EMVA analyses have multiple dimensions of data. (All of the CA’s executed in this work have between 24 and 26²⁰ dimensions but most of that information is “compressed” into the first three “mappable” dimensions.) Thus, any two-dimensional projection necessarily fails to

¹⁹ Because the X (horizontal) axis in both projections is dimension one and we are rotating around that dimension, a vertical line can be drawn from Figure 1A down through Figure 1B and each of the 27 individual texts in the GNT will align vertically

²⁰ The minimal number of dimensions in EMVA methods is $1 - N$, where N = the row or columns in the contingency tables, whichever is least.

represent information in the third dimension. Accordingly, we used a proprietary data reduction tool, PositionSolve™ to rotate the CA so that the six texts with positive component loadings for the first component (see Figures 8.1A and 8.1B) lay upon the plane of the page.²¹ This yields a truer perspective of the relative distances between these NT texts. The rotated perspective demonstrated that Acts, John, and Revelation lie relatively far apart from each other while the Synoptic Gospels lie relatively close to one another. The Euclidean distances between the texts pairs are as follows:

	<i>Book Pairs</i>		<i>Euclidean Distances Between Texts Adjusted by Inertia (First Three Dimensions)²²</i>
1	Luke	Matthew	2.29
2	Matthew	Mark	3.17
3	Luke	Mark	4.84
4	Luke	Acts	10.87
5	Synoptics	Acts	12.75
6	Luke	Revelation	16.42
7	Synoptics	Revelation	16.55
8	Synoptics	John	19.02
9	Acts	Revelation	19.09
10	Luke	John	19.87
11	Acts	John	27.19
12	John	Revelation	31.93

Table 8.1: Generalized Euclidean Distance²³ between Selected Pairs of NT Texts

Though the primary purpose of this section is interpretive, it is appropriate here to introduce a finding that appears repeatedly in this work. Both Figure 8.2 and Table 8.1 problematize the *de facto* historic view that NT texts cluster primarily by authorship. Two consistent observations are important in this regard.

²¹ PositionSolve™ was provided courtesy of Decision Support Sciences.

²² While, technically, a chi-squared distance is indicated here, given the modestly similar mass of these books and the pedagogical nature of our effort, Euclidean distance serves as a close approximation. The cluster centers were, however, calculated using chi-squared distance.

²³ This distance was developed from the column scores of a CA in which the columns were the texts of the NT and the rows were the top 100 most variant lemmas (via AAVASR, the average absolute value of the adjusted standardized residual). To obtain Table 1 Euclidean distance was multiplied by the inertia of each dimension to adjust for the information content in each dimension, and then by 100 for easier visual comparison.

- Using lemmas (and in fact across all measures at the rank of lexis) the Synoptic Gospels are quite tightly grouped. If the primary component of this data reflects authorship, this is *not* what one would expect given that the Matthean/Markan/Lukan authors (or editors or redactors) are universally assumed to be separate individuals. If one should object that the close proximity of the Synoptics is due to common source material (such as Q), then one would expect Mark to be closer to Matthew and Luke, and that unique Matthean material (M) and unique Lukan material (L) would place these two Gospels at some relief from one another. Actually the opposite is the case, per Table 8.1. A more perspicuous explanation for the observed text clusters at the rank of lexis is that the Synoptic Gospels group together because they constitute a synoptic *genre* (i.e. Greco-Roman biography or the like).²⁴
- Moreover, per Figure 8.2, the texts of Luke and Acts lie relatively *far* from one another. Given that NT scholarship almost without exception confirms that Luke and Acts were written entirely or substantially by the same author, this is also not what one would expect if the primary component of the variance structure of the GNT reflects authorship. If, on the other hand, genre were the primary driving factor in the conformation of these texts, and authorship a more secondary factor, then one would expect Luke and Acts to be located at some distance from one another, and for Luke to be closer to Acts than the other two Synoptics. This indeed is the case.

²⁴ We propose that the greater similarity between Luke and Matthew reflects the greater similarity in lexical stock between M and L. That is, Mark stands apart because his non-common Marcan material (or his redaction of Q) has less in common lexically with M and L than M and L have with each other.

8.4 Visualizing Lexematic, Paradigmatic and Semantic Data Using CA

8.4.1 Visualizing the Lexematic Data

As our first example in this section we explore lemmas, the canonical lexeme, as follows:

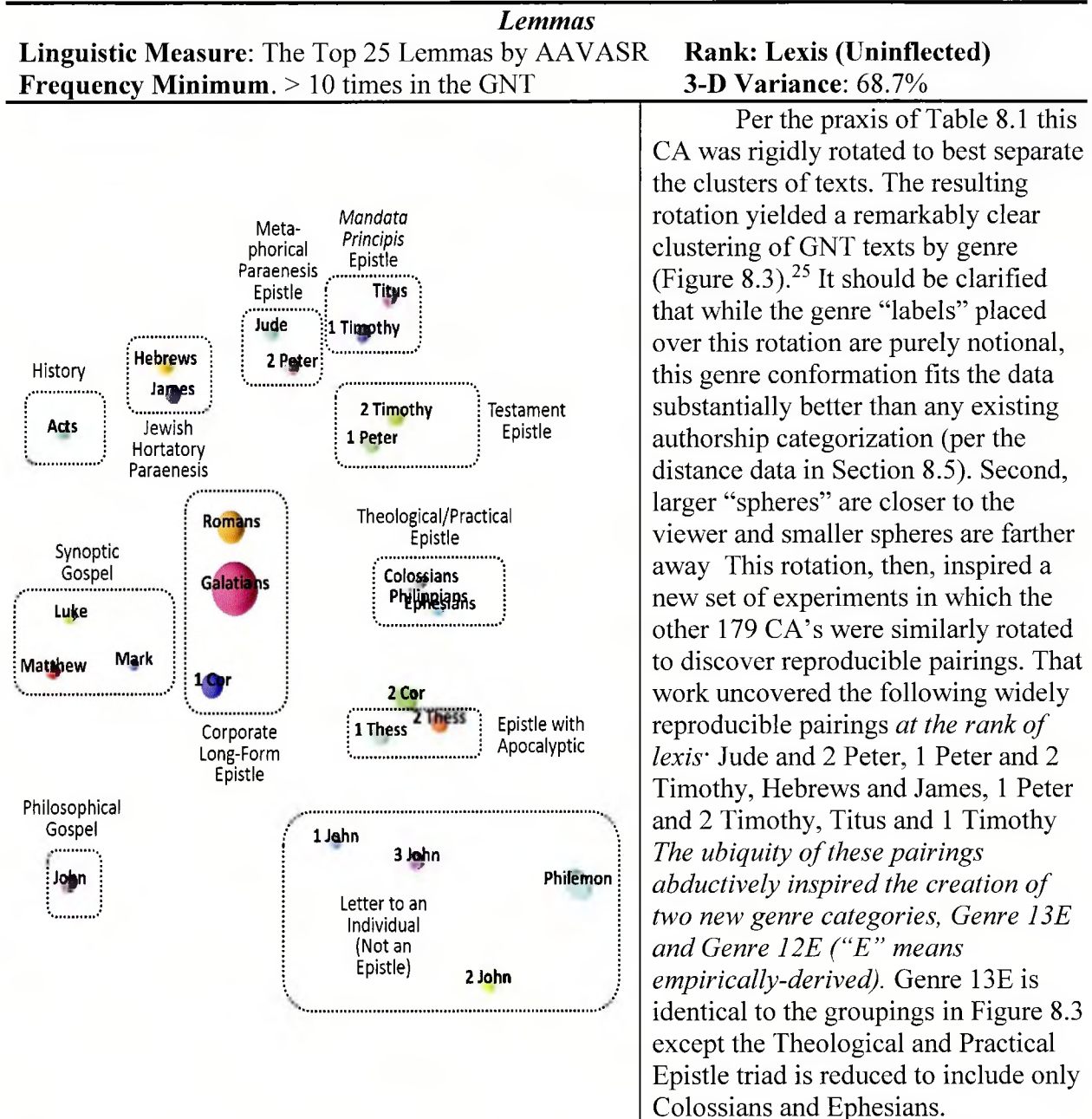


Figure 8.3

²⁵ This association exists for all other measures at the rank of lexis (i.e. for inflected words, and semantic domains).

8.4.2 Visualizing Paradigmatic Data

As a second example a CA of one of the three proxies for paradigmatics is presented. In this case the largely traditional grammar-based AGNT annotation was chosen.

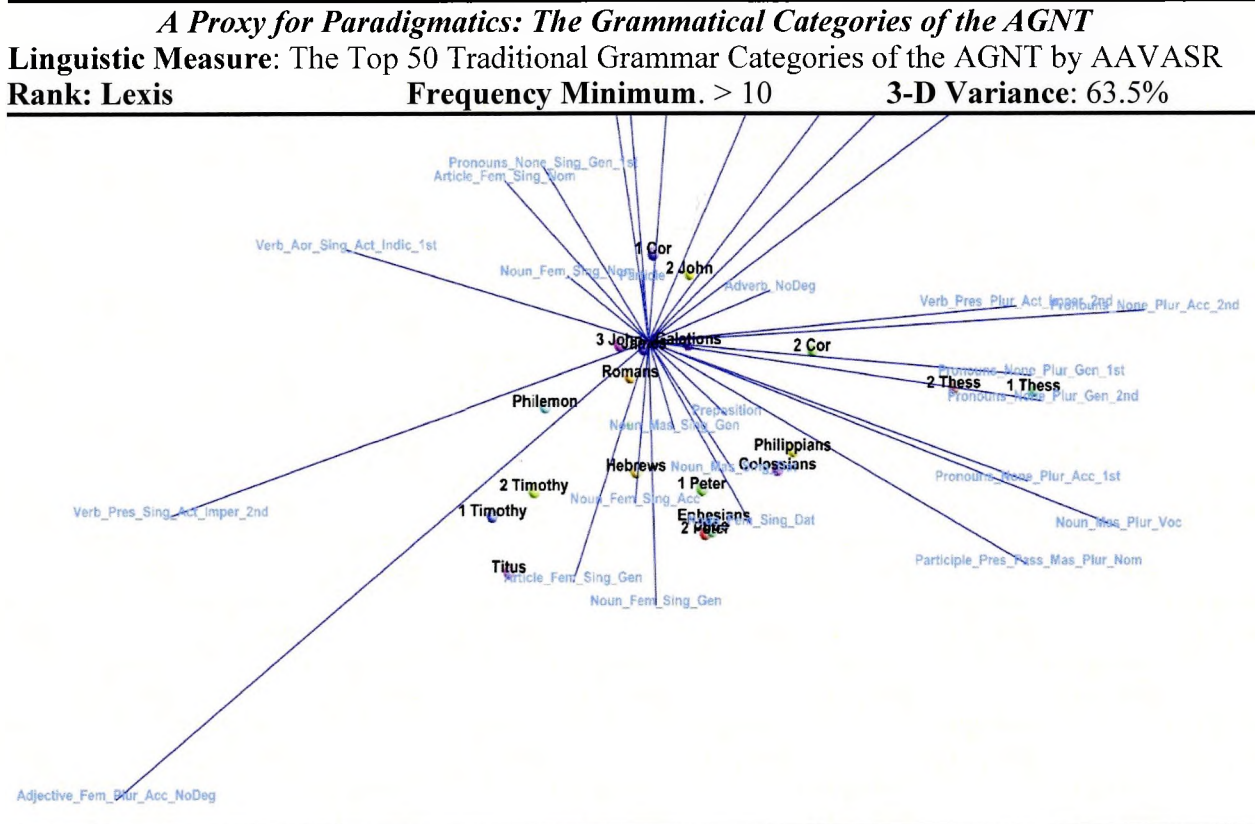


Figure 8.4.

Two findings in Figure 8.4 command our attention.

- 1) As was observed with lemmas and inflected lexemes, both Thessalonian texts are significantly displaced from the rest of the epistles. This appears to be the first *paradigmatic* confirmation supporting Mealand’s initial discovery of the pairing of 1 Thessalonians with 2 Thessalonians.²⁶

²⁶ See Mealand, “The Extent of the Pauline Corpus,” 86. In our view Mealand’s judgment may prove right even though it was premature because (a) Mealand did not recognize the extent to which the overall clustering of the GNT texts is by genre rather than authorship, (b) he used only 19 measures rather than the 50 we use here and (c) his measures were developed apart from any formal feature set selection approach. Intriguingly, substantial aspects of Pitts’ sociolinguistically predicated profile for the Pauline corpus are validated here. Specifically, the CA data lends clear empirical support to his second register profile (the Thessalonian pairing), modestly less support for the third

- 2) The distance of the Pastoral Epistles from the rest of the epistles can be seen to be due to the high frequency of feminine plural accusative adjectives and present active imperatives (second singular). Rotation demonstrates that the displacement of the Pastoral Epistles is slightly closer to the map center than that of the Thessalonian pairing.

profile (the *Hauptbriefe*), strong support for the fourth (especially in terms of the Colossians and Ephesians pairing), and excellent support for profile five (the Pastoral Epistles) per Pitts, "Style and Pseudonymity in Pauline Scholarship," 136–37.

8.4.3 Visualizing Semantic Data²⁷

In this third CA example, a different linguistic strata of the linguistic data of the GNT is projected, the strata of semantics.

Linguistic Measure: The Top 50 Major Semantic Domains (Disambiguated) by AAVASR
Rank: Lexis **Frequency Minimum.** > 10 **3-D Variance:** 69.8%

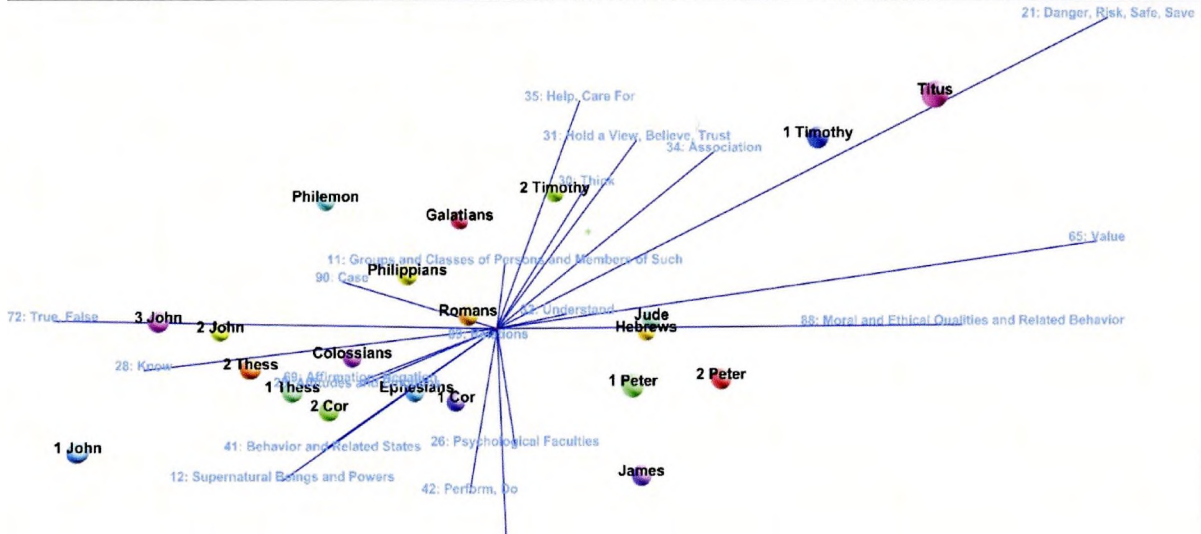


Figure 8.5A.

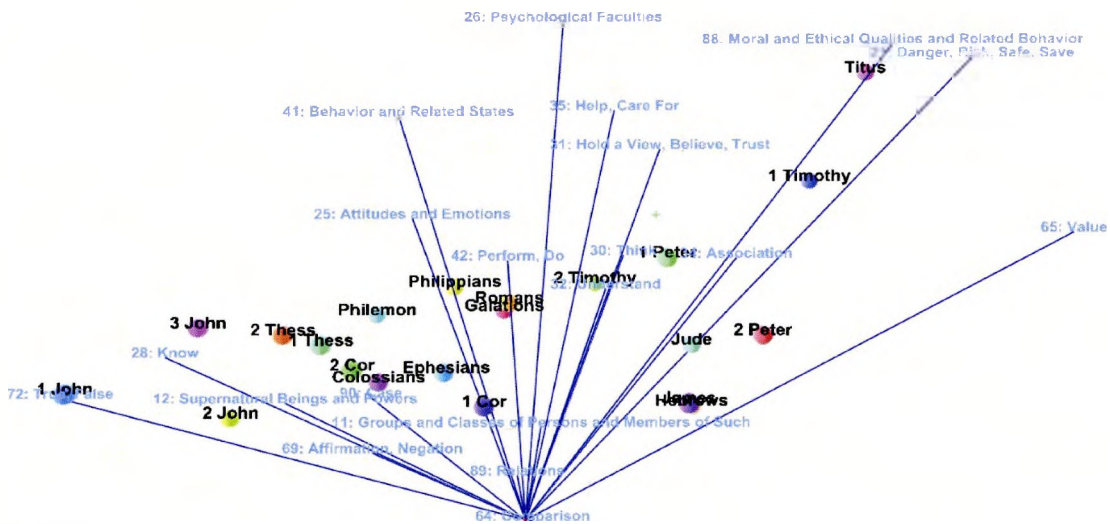


Figure 8.5B.

²⁷ Nondisambiguated major and minor semantic domains have been directly imported from Louw and Nida's Greek-English Lexicon (courtesy of the OpenText.org database). We gratefully acknowledge the assistance of Rick Brannan of Logos Bible Software for providing us with the disambiguated Louw and Nida domains.

In Figures 8.5A and 8.5B the biplots were decluttered by focusing only on the epistles and by displaying only 15 of the top 50 major domains. In addition, per the praxis of Figure 8.2, the CA projection of Figure 8.5A was rotated upwards around the horizontal axis to yield the projection in Figure 8.5B. These actions yielded two clear findings:

- 1) Similar to lemmas, disambiguated semantic major domains seem to map remarkably well to notional constructs of genre. (Compare 8.5A and 8.5B to Figure 8.3.)
- 2) Semantically, per Figures 12A and 12B, the Pastoral Epistles are deviated from the main body of the rest of the epistles by their distinct semantic content (major domains 21, 65 and 88).

8.4.4 Summarizing the CA Data Lexically, Paradigmatically and Semantically

This brief excursion into lemmas, paradigms, and semantics as analyzed by CA demonstrates, by example, what our larger suite of 180 CA runs also demonstrated:

- 1) Although the CA projections analyzed a wide range of linguistic measures across the different ranks, systems, and strata of the GNT, and while their book conformations were, as expected, not identical given that they encode different kinds of linguistic information, there are large morphological similarities across virtually all the CA runs. These included: (a) the large separation between the epistles, Revelation, Acts, and Gospel of John, (b) the diffuse cluster of the smaller Johannines, (c) the tight clustering of the Synoptic Gospels and (d) the diffuse clustering of the epistles.
- 2) At the lower spans of rank, especially with lemmas and inflected lemmas (words) reproducible clusters of GNT exist such as those in Figure 8.3. Notable among these was the pairing of the Thessalonian letters, the pairing of Ephesians and Colossians, the pairing of Jude and 2 Peter, and the pairing of 1 Timothy and Titus.

- 3) The *Haupbriefe*, while always within the same quadrant, seldom formed a convincing cluster, even a diffuse one.

The overwhelming impression is that the GNT texts clustered by genre rather than by authorship. This qualitative impression is quantitatively confirmed in our next analysis.

8.5 Using Euclidean Distance to Test How Well The Theories Fit the CA Data

The overall *qualitative* impression yielded by the CA data just presented is that the texts of the NT cluster by genre rather than by authorship. This raises the question as to whether there is a way to confirm this finding *quantitatively*. In the first attempt to do so all 180 CA's were rotated to qualitatively inspect each for a genre-like or authorship-like text clustering. Next these same CA's were rotated in the fourth through the sixth dimensions. In both cases the intention was to discover an "authorship-like" clustering of texts per the work of Burrows, Forsyth, Juola, Baayen, and others.²⁸ The sheer drudgery of that work made it clear that a more objective and efficient way forward was needed to identify which authorship and genre theories best fit the data. Accordingly, Euclidean distance spread (EDS) was chosen for this task. As a simple distance-based measure, EDS tests how well the text categories of the various theories (whether they be authorship or genre theories) matched the actual locations of the texts revealed by the CA analyses. A further explanation of EDS is in order. *Definitionally*, EDS is arguably the simplest distance-based metric for determining how well a given theory (such as an authorship or genre theory) fits the CA data. *Mathematically*, EDS is calculated by measuring the Euclidean (simple) distance between the members of a given cluster and the center of that cluster. *Interpretively*, a large spread indicates either a diffuse cluster or that one or more elements do not "fit" within that

²⁸ It is typical for the first component extracted by extractive multivariate analysis (EMVA) in a mixed genre corpus to be associated with genre rather than authorship. See Burrows, "The Statistical Analysis of Narrative Style," 64; Forsyth et al., "Investigating the Authenticity of the Consolatio," 383; Juola and Baayen, "Authorship Identification by Cross-Entropy," 63. Cf. Burrows, "The Interpretative Nexus between Analysis and Information," 92–102.

cluster. The power of any distance-based spread measure, is that any theory can be tested using the same CA map. Two methods of calculating the EDS, one normed and one not normed, are presented in Figures 8.6 and 8.7 below

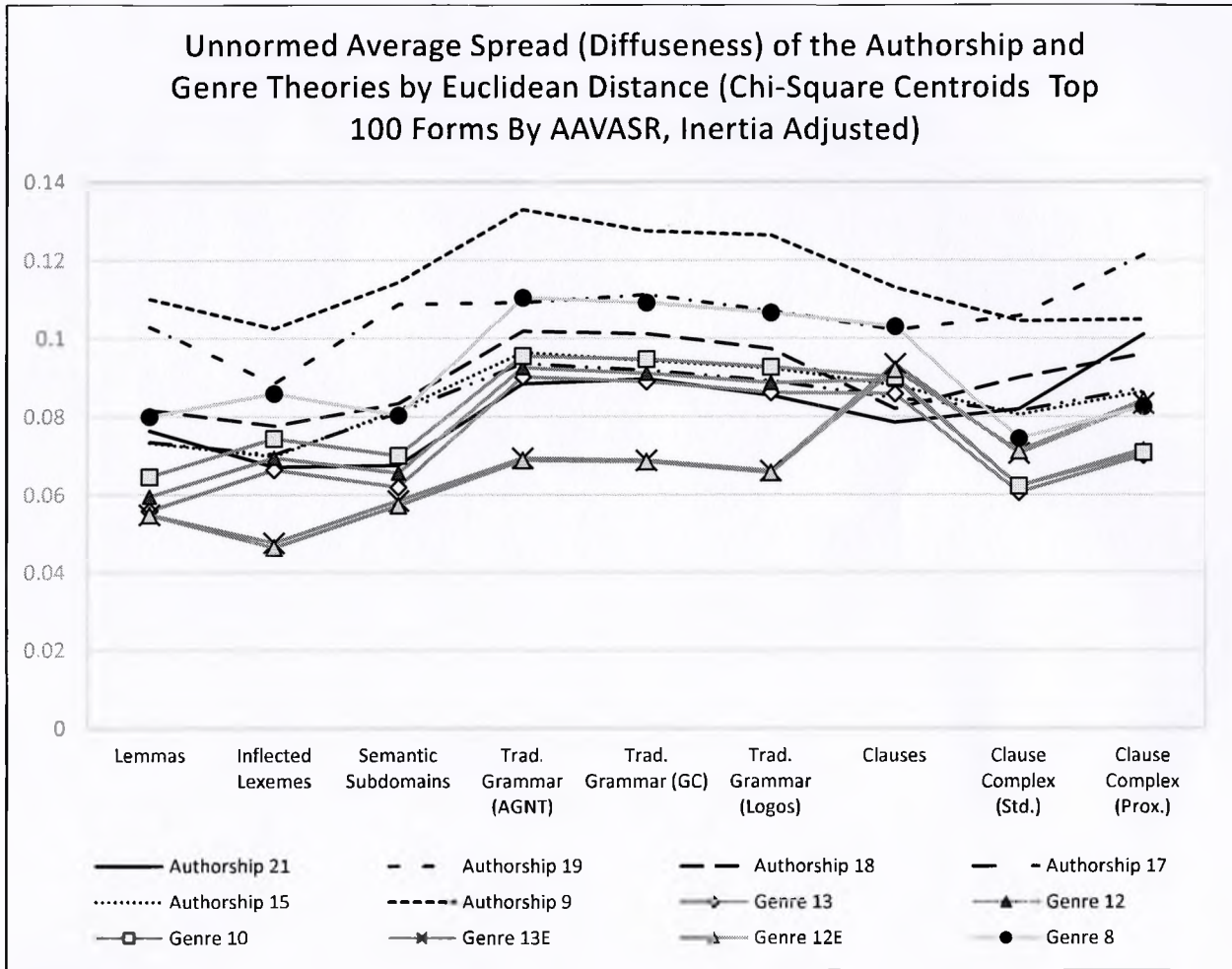


Figure 8.6

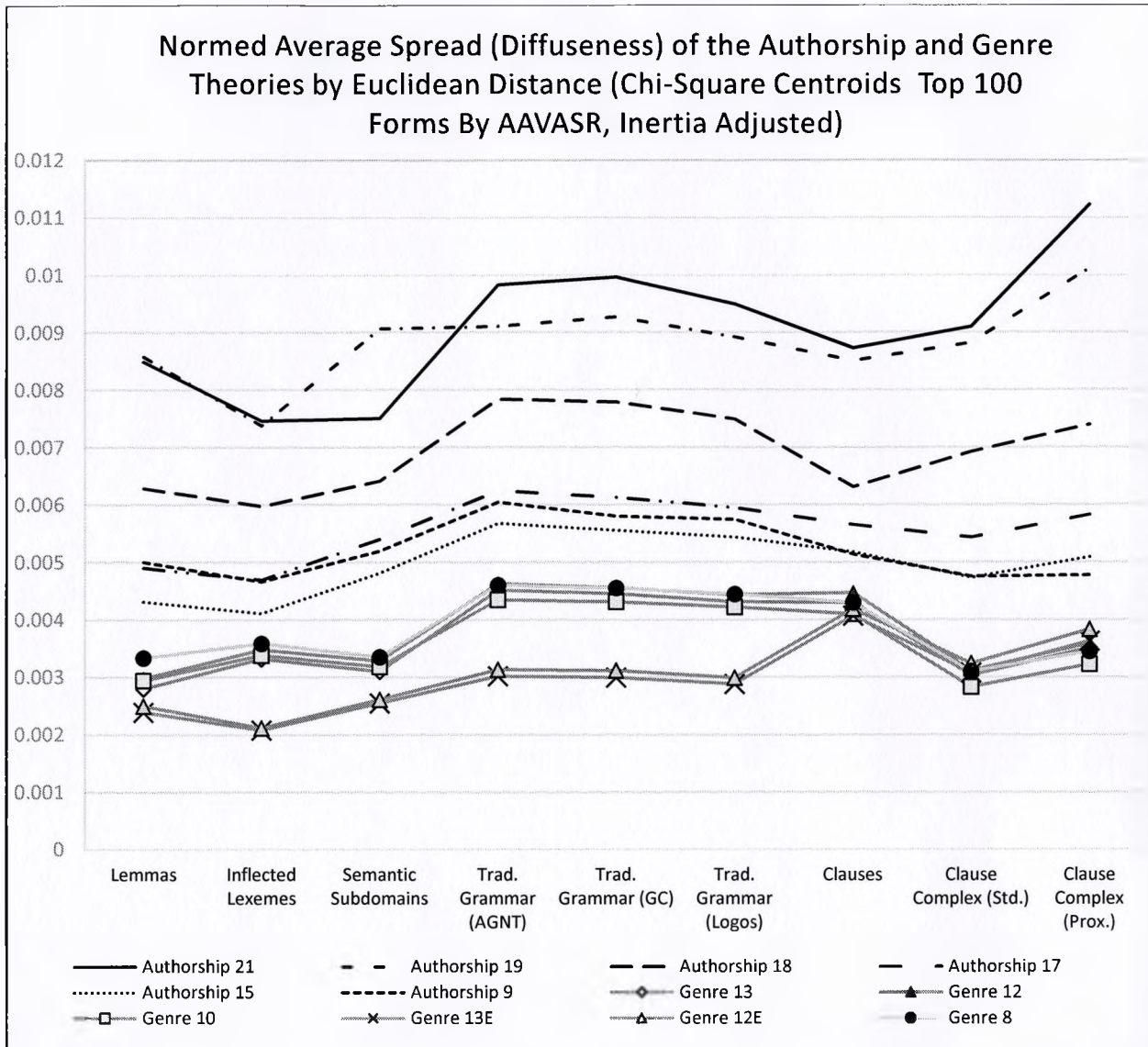


Figure 8.7

Discussion of the EDS Findings

Figure 8.6 presents the non-normed EDS data, and Figure 8.7 is the normed EDS data.

Norming the data is indicated because it is topologically unfair to compare a theory that is required to fit only a small number of texts (e.g. 9 texts for the Baur 21 authorship theory) to a theory that is required to fit far more (e.g. 21 texts for traditional authorship theory). While the mathematics of how to calculate the norming adjustment between Figures 8.6 and 8.7 is surprisingly difficult, the determination was made to simply divide the non-normed data from

Figure 8.6 by the number of texts that actually reside within its clusters (a cluster is defined as a grouping of two or more texts). This then yielded Figure 8.7. This approximation is certainly adequate given that even without norming, the non-normed data demonstrates many of the same findings. Four findings are noteworthy: (1) All theories of genre in Figure 8.7 and most in Figure 8.6 display less spread (higher fit across all nine linguistic measures tested) than all theories of authorship. (2) The empirically derived genre theories (13E and 12E) outperform all other genre theories in terms of fit to the CA data. (3) For syntagm lengths above the rank of the clause (clause complexes) genre theories still outperform the authorship theories (Figure 6) but become less distinguishable from one another.²⁹ (4) Even when the data is non-normed, per Figure 8.6, genre displays less spread (fits the data better) for all linguistic measures except at the rank of the clause. In conclusion, then, across all linguistic measures tested, whether measured qualitatively or quantitatively, or whether extracted by AAVASR, APASR or by frequency, the variation seen in the GNT is disproportionately generic rather than authorial in nature.

²⁹ This indistinguishability may relate to the fact that measures at or above the rank of the clause are much more highly skewed—they contain a higher number of low-frequency forms.

9 Results Part IV: Exploring GNT Information Loss Via Information Theory

This chapter was unplanned. As such I could have placed it in the Appendix, because in essence the entire argument necessary for the conclusions to be summarized in chapter ten already exist. Yet, I have decided to promote this chapter out of the appendix because, more than any other chapter of this work, it exemplifies the benefit of an abductive design. Perhaps, then, a more whimsical title for this chapter could have been, “How Abduction Makes Us Follow Bread Crumbs: Information Loss Can Be Our Gain in Understanding Genre and Authorship in the GNT.” Accordingly this chapter came about because I sought to discover more about information loss in both the genre and authorship theories in the GNT. It seemed, in short, that univariate (or more properly bivariate) information theory (IT) had much more that it could deliver in contributing to such an understanding beyond what it did for us in Section 6.4. To begin our abductive exploration, recall from Chapters Six and Seven that the traditional theory of authorship (a) demonstrated the highest overall amount of *association* between language among all the authorship theories, but (b) the lowest *uncertainty coefficient* (UC). It became apparent that either information was being disproportionately destroyed in the traditional authorship grouping or something else was causing the UC to be so low. The answer we argued for in Section 7.6 *seems* compelling; the differential loss of information in the traditional authorship theory was hypothesized to be due to *equivocation*. That is, the high information content, low-frequency instances of language associated with genre were being lumped together. The most signal example of this is Paul, who according to the traditional author theory wrote 13 texts of the NT across six genres. We then argued that by mixing all these genres together we substantially confounded (mixed together) the specialized genre-specific information of those

texts. This effect could hardly be trivial given that the same chapter (Section 7.4) demonstrated that genre disentangled from authorship (G|A) was uniformly stronger than authorship.

So why this chapter? Overall the argument as laid out above seems reasonable—but it lacks a smoking gun. Thus, in this admittedly exploratory chapter the goal is to find that smoking gun. The approach will follow the same path taken in chapters six through eight. First, I will provide some necessary history on the prior intersections of information theory, linguistics, and computational stylistics, then provide a short primer, and finally explore discrete steps in answering questions related to information loss and the GNT. A disclaimer is due here. This work is very new, and while the data is robust its interpretation requires further experiments and perhaps other notional theoretical frameworks. These will be summarized in the concluding chapter.

9.1 The History of Information Theory and Linguistics

In proposing an information theoretic approach to the language of the NT, we must take careful stock of the checkered past of information theory. The earliest intersection of these disciplines ended, to speak bluntly, in abject failure. The story of Information Theory and its implementation in linguistics is the stuff of legend. It involved narrow escapes from Nazi terror, top secret wartime science, Rockefeller money, the cold war, secret fact-finding missions, political chicanery, an “A” list of the leading academics worldwide, a theory of just about everything, MIT, and an unknown post-doc named...Noam Chomsky. The end result of this first “information theory attached to language” peaked at the high-water mark of twentieth-century structuralism, and in fact assisted in its collapse, so one might ask, why attempt to resurrect it?

Three reasons compel us. *First*, though initially poorly appropriated by linguists, Shannon’s original theory and the field of information theory that sprang from it has been enormously successful. It has yielded immediate and wide application in fields as disparate as

literary theory (prose readability),¹ aesthetics,² communication,³ physics,⁴ statistical inference,⁵ digital compression,⁶ neurobiology⁷ and data mining.⁸ It is clear, then, that Shannon's theory works—when indeed the system in question is truly a single channel information system. *Second*, and this is directly related to the prior point, there is more than ample evidence now that appropriate domain-level applications in language surely *do* exist, because IT is currently experiencing a clear renaissance in linguistics.⁹ *Third*, given the prior two points, what if the initial problem with the application of IT to language lay not with IT theory *per se*, but with identifying its appropriate domain-space within language? That is, what if Jakobson was actually correct that natural language (or at least components of it) can be modeled with information theory, but the problem lay in the fact that Shannon's application operated at only one of perhaps dozens (even hundreds?) of ways language summatively transmits its semantic freight?¹⁰

¹ Taylor, "Recent Developments in the Use of "Cloze Procedure"," passim; Taylor, "Cloze Procedure," passim.

² Moles, *Information Theory and Esthetic Perception*; Berlyne, *Aesthetics and Psychobiology*.

³ Rogers and Kincaid, *Communication Networks*.

⁴ Dewar, "Information Theory Explanation of the Fluctuation Theorem," passim.

⁵ Kullback, *Information Theory and Statistics*; Akaike, "Prediction and Entropy," 387–410; Rao, *Linear Statistical Inference*.

⁶ Wallace, "The JPEG Still Picture Compression Standard," passim.

⁷ Jaynes, "How Does the Brain Do Plausible Reasoning?," passim

⁸ The applications of IT to data mining can be classed into two categories: quantifying the information content for further data analysis (Kantardzic, *Data Mining*, 68–70), and structural modeling (Emmert-Streib, *Information Theory and Statistical Learning*, 183–207; Krippendorff, *Information Theory*, 32–85).

⁹ In Pereira's words, "Today, after many years on the defensive, the information-theoretic approach is again thriving and has led to practical successes in speech recognition, information retrieval, and, increasingly, in language analysis and machine translation." Pereira, "Formal Grammar and Information Theory," 1240. For examples see especially Thoiron, "Diversity Index and Entropy" 198–200, and Arppe, "Univariate, Bivariate, and Multivariate Methods in Corpus-Based Lexicography," 154–57. For an early articulation of the utility of information theoretic approaches see Harris, *Structural Linguistics*. Subsequently Harris, who was Chomsky's mentor, articulates four constraints for using IT in the service of linguistics (Harris, *Language and Information* as quoted in Pereira, "Formal Grammar and Information Theory," 1240–41).

¹⁰ Perhaps most notable in the more recent works are efforts in probabilistic grammar that demonstrate the inadequacy of the Chomskyan rubric—just as Chomsky did to IT-based approaches years before. Notable among these are Ellis, amended by Kornai, who demonstrated that probabilistic languages (p-languages) exist that are not generated by probabilistic finite state grammars (PFSG). Kornai memorably concludes, "the variety of probabilistic structures cannot be reduced to the variety of algebraic structures; there are more things in probabilistic heaven and earth than are dreamt of in your grammatical philosophy" (Kornai, "Probabilistic Grammars and Languages," 319). In Vaux's view, Jakobson, in league with all other formalists at that time, was attracted to information theory in the belief that language needed an efficient storage system due to the lack of memory storage in the brain. Consequently, Jakobson underspecified language (Vaux, "Syllabification in Armenian," 118.) Cf. also with Sankoff

Specifically, what if natural language is not a single channel, single phase, lineal data stream as required by Shannon's instantiation of Markov's theory? What if instead, as we have already asserted, language is a broad, multistratal, multirank and multisystemic complex that communicates through multiple channels simultaneously? Such a state (no pun intended) would change the information value of each morpheme or lexeme in the realized lineal data stream. This, in fact, is exactly what Sgouros has indirectly discovered in repeated Cloze reading experiments.¹¹

Motivated by this as a possibility, and certainly consistent with the SFL view that language is a complex rather than a simple social semiotic, a deeper exploration of other IT measures seemed indicated. This line of thinking also raised the possibility of studying language phenomena in multiple channels *and* comparing that to structural equation modeling insights that assist in discovering latent layers of information that may *interact* with one another.¹² In this scenario the actual information value in each data channel may be able to be modified or quantitatively changed by concurrent data streams flowing in separate channels. To make matters even more complex, over the course of a text's lineal "reading" those original channels may intersect with information streams from different authors/editors/redactors. Given these kind of considerations, the opportunities seemed rich in digging into this phenomenon. The immediate question then became: What would such an architecture look like when applied to the NT? Three implications seemed to follow:

who views Jakobson's effort as an easily dispatched strawman because it was probabilistically naïve (Sankoff, "Probability and Linguistic Variation," 217).

¹¹ Sgouros, "What Is Context for?," *passim*.

¹² If our optimism is justified, then it may be possible to say of the mid-twentieth century experiment with information theory that seldom in the history of academic discourse has the mainstream of a discipline so steeped in a construct (linguistic semiotics) so mishandled a discovery so thoroughly central to its own propositions.

First, if language does indeed communicate information through multiple channels simultaneously, an adequate measurement model must be developed to transduce these multiple channels of information.¹³

Second, because information theory introduces into linguistics not only the channel but the factors that inhibit information from being received unambiguously (equivocation and noise), we must quantify the relation between noise and equivocation and seek their sources in language.

Third, information theory, especially in its more mature incarnations, provides ways in which predictive and causal questions can be addressed. These involve building structural models of language “upwards” from the corpora of language itself. Because nothing prevents information theory-based structural models from being run in parallel with more traditional structural modeling (e.g. LISREL and the like), information theory can be used to corroborate or nuance traditional structural modeling. Such an approach should be executed to provide multiple, parallel insights into linguistic analysis, description, prediction, and causation.

9.1.1 Information Theory and the New Testament: Some Baby Steps

Although I have described a broad domain of usefulness for IT theory directly above, the primitives, the simple bivariate measures of information theory, are also quite appropriate for addressing far more modest questions in language analysis as well. Three such questions are related to our muse of genre and authorship in the GNT:

- Which authorship or genre theories communicate the most information? Why?
- Which language fraction encodes the most information? What does this mean for genre and authorship in the GNT?
- Does TI loss differ by language measure in the GNT?

¹³ Cf. Rulon’s postmortem on descriptivism written in 1963: “The levels-idea could have been - indeed, still can be – thought out much further than it was; for all that was done with it, it remains largely a neglected possibility in twentieth-century structuralism.” Wells, “Some Neglected Opportunities,” 44.

To achieve these ends, however, requires a minimal working understanding of information theory via the short primer I provide below.

9.2 Studying the GNT Using Information Theory: A Primer

Fortunately, a working understanding of IT and its interpretation can be had by grasping a very modest set of general and specific propositions.¹⁴ Before enumerating these propositions, however, two delimiting points are needed. First, consistent with the rest of the analytics in this study, our exploration will focus on the analysis of contingency tables. Second, by convention, the categories of the nine language measures constitute the rows of the table and the NT text groupings (the authorship or genre theories) constitute the columns. With these definitions, I can now propose the necessary general propositions. Ten are relevant:¹⁵

- GP₁: A sender (S) sends a message (M) encoded in language. Encoding exists at all language strata and at all ranks of language.
- GP₂: There are three measures of *observed variation* or *entropy* in a contingency table. The sender's entropy is the observed variation in the rows, H(S), the receiver's entropy, is the observed variation in the columns, H(R), and the joint entropy is the observed variation in the row by column table H(SR).¹⁶
- GP₃: Equivocation (E) is that component of the sender's entropy, H(S), in which the receiver (R) is unable to differentiate between two messages. As ambiguous communication, E communicates zero information content to the receiver (R).

¹⁴ In addition to answering these questions, IT analysis provides a complementary insight into the language of the GNT in two ways. First, IT theory provides unique measures not paralleled by Fisherian statistics (information content, terseness, separability, and the like). Second, just as HLLA provides a mathematical model which can disentangle main effects (e.g. authorship from genre), IT can do the same, providing a complementary way to understand the relationship between authorship and genre.

¹⁵ For a more detailed but still accessible summary of the separate elements of information theory, see Cover et al., "Elements of Information Theory," passim.

¹⁶ Specifically, entropy is the total logical information content in a message minus the uncertainty such a classification loses. Hence it is the remaining observed variety in a message. That variety has two further components, information and uncertainty.

$$\text{Eq 1: Entropy} = \log_2 n - \sum_{a \in A} \frac{n_a}{n} (\log_2 n_a) \quad \text{Where: } n = \text{total information in a message, } n_a \text{ is the information once it is grouped for any reason.}$$

This reduces to the equivalent and more familiar original notation of Shannon (Shannon & Weaver, 1949)

$$\text{Eq 2: Entropy} = - \sum_{a \in A} p_a \log_2 p_a \quad \text{Where: } p_a = n_a / n; = \text{the probability of } n_a \text{ with respect to } n.$$

- GP₄: The total information content communicated from the sender (S) to the receiver (R) is termed TI:(SR). This TI:(SR) is that component of H(S) communicated unambiguously to the receiver (R).
- GP₅: The total communicated information content is also the sum of the sender's and receiver's entropies minus the joint entropy: TI:(SR) = H(S) + H(R) – H(SR).
- GP₆: The sender's entropy is the sum of the total information plus equivocation: H(S) = TI:(SR) + E.
- GP₇: Noise (N) is that component of the receiver's entropy H(R) in which the channel or the environment adds ambiguous information to the total information of the message TI:(SR). As unambiguous communication, noise (N) adds zero information content to the message received by the receiver (R).
- GP₈: The receiver's entropy H(R) is the observed variation seen in columns. It is the sum of the total information and noise: H(R) = TI:(SR) + N.
- GP₉: The percent of the sender's classification able to be known from the receiver's categories = UC (S) = TI:(SR) / H(S). This is the *row-dependent* uncertainty coefficient.
- GP₁₀: The percent of the receiver's classification able to be known from the sender's message categories = UC (R) = TI:(SR) / H(R).¹⁷ This is the *column-dependent* uncertainty coefficient.

These collective general propositions can be visualized by adapting Krippendorff's storied schematic of Information Theory (see Figure 9.1).

¹⁷ This can be seen first by comparing the equations of the two PRE measures: First the algebraic equations for Goodman and Kruskal's tau:

$$Eq\ 1: \tau(R|C) = \frac{n \sum_{i=1}^r \sum_{j=1}^c \frac{n_{ij}^2}{n_i} - \sum_{j=1}^c n_j^2}{n^2 - \sum_{j=1}^c n_j^2} \qquad Eq\ 2: UC(R|C) = \frac{\sum_{i=1}^r \sum_{j=1}^c n_{ij} \ln \frac{nn_{ij}}{n_i n_j}}{\sum_{j=1}^c n_j \ln \frac{n_j}{n}}$$

Where: n_i and n_j are the row totals for row i and the column totals for column j , respectively, n is the total frequency of the tables, n_{ij} is the frequency within the i,j th cell, r is the maximum columns, and c is the maximum rows.

Per equation 1, UC weights low-frequency information far more highly than Cramér's V or G&K's tau because it takes the natural log of the cell contributions. G&K's tau, in contrast, similar to all variance-based metrics, calculates variance on the sum of squares. Demonstrating this heritage, tau yields profiles similar to chi-square measures such as Cramér's V. G&K's tau, in fact, is equal to R^2 for contingency tables (Margolin & Light, 1974). Accordingly, Liebetrau notes that this yields another interpretation for G&K's tau, the proportion of variance in the column attributable to the row.

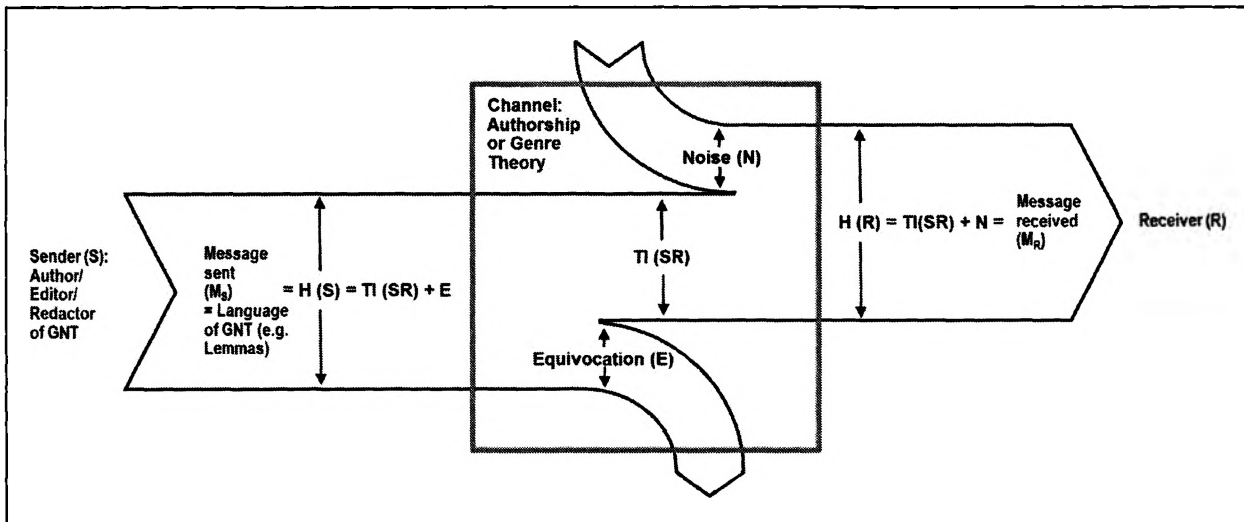


Figure 9.1. Information Theory Applied to the Language of the GNT (as adapted from Krippendorff)¹⁸

In our specific case in which all $H(S)$ values in a comparison set are identical (for the reasons set out below), information theory allows us to further define the message (M) as any combination of linguistic measures from a sender (S) (or author/editor/redactor) to any real or hypothesized set of receivers. Doing so does not require that we conceive of the authorship or genre groupings as the receivers, but rather as the channel (the “pipeline”) through which the communication proceeds to the receivers. Unique to our experimental design, for a given language measure all the rows are identical (e.g. 138,019 instances of lemmas exist in the GNT in 5,413 categories) but the columns (the various authorship and genre theories) differ. Given these qualifications, ten additional specific propositions follow:

- SP_1 : Each message is a single instance of the language measure being explored (e.g. one of 138,019 instances of lemmas in the GNT in 5,413 categories).
- SP_2 : A column classification is correct when the instance of a message (from a language measure row) is correctly classified into the correct column (genre or authorship) category.
- SP_3 : The only column classification that can be directly confirmed to be correct is the control, where that form is known to be classified into one NT text.
- SP_4 : Although correct classification of the competing authorship and genre theories cannot be objectively achieved (per SP_3), the various theories can be compared objectively using a

¹⁸ See Krippendorff, *Information Theory*, 25.

number of metrics that *are* meaningful for NT interpretation. Here we will focus on two of those metrics: TI:(SR) and UC(R).

- SP₅: Given that all contingency tables have the same number of rows (e.g. 5,413 categories of lemmas) and the same number of observations (the 138,019 lexemes of the GNT) all authorship and genre theories have the same sender's entropy, H(S).
- SP₆: Given that all theories have the same H(S), and given GP₅, $TI:(SR) = H(S) + H(R) - H(SR)$, then the quantitative difference in TI:(SR) can arise only from the difference in the receiver's entropy H(R), and the joint entropy H(SR). Stated equivalently, TI:(SR) can only change due to a change in the main effect of the column, H(R), or the interaction between the row and the column, H(SR).
- SP₇: Given that all theories have the same H(S), and given GP₅, $TI:(SR) = H(S) + H(R) - H(SR)$ and given GP₉, $UC(S) = TI:(SR) / H(S)$, then UC(S) can change only due to the main effect of the column H(R) and the interaction between the row and the column, H(SR).
- SP₈: Given that all theories have the same H(S), and given GP₅, $TI:(SR) = H(S) + H(R) - H(SR)$, and GP₁₀, $UC(R) = TI:(SR) / H(R)$, then UC(R) can change only due to the main effect of the column H(R), and the interaction between the row and the column H(SR).
- SP₉: Given that all theories have the same H(S), and given GP₉, $UC(S) = TI:(SR) / H(S)$, TI:(SR) is completely collinear to UC(S).¹⁹ (The shape of the profile is identical.)
- SP₁₀ When all *hapax legomena* (forms which occur only once in the NT) are analyzed as a group, H(S) and H(R) are maximal: $TI:(SR) = H(R)$, $N=0$, and $H(S) = H(SR) = H(R)+E$.

This brief primer on information theory closes with a few final comments related to interpreting information theory-based measures:

- Entropy-based measures from information theory cut to the core of the linguistic argument for authorship, an argument which reasonably assumes that there are unique patterns to an author's style. Said in terms of IT, those unique patterns that are otherwise in low frequency in the corpus, have high information content; they are "surprising"²⁰ and can't be predicted. Information theory "notices" these surprises.
- Combining the assumption of pseudepigraphy with information theory, the following can be inferred: If (a) nothing else contributes substantially to creating unique patterns in texts and (b) the receiver's grouping categories are standardized to allow information to be compared

¹⁹ Collinearly means that the each element in a resultant vector A (a vector is just a sequence of values) is equal to the values in a source vector B multiplied by a non-zero constant. In our case this means that the ratio between the elements of vector A is identical to the ratio of the corresponding elements in vector B.

²⁰ See Krippendorff, *Information Theory*, 14.

across theories without a grouping bias, then (c) a corpus with unique authors who have unique patterns of communication will have the highest information content.

- As a corollary to the above, if another main effect or interaction effect contributes substantially to creating these unique patterns (e.g. genre), a corpus with unique authors who have unique patterns of communication may *not* have the highest information content.
- Although closely associated with markedness theory, information theory is not identical to it.²¹ Markedness theory asserts that infrequent linguistic items are more “marked” and allow the analysis of language to be understood in terms of binary oppositions. Information theory, however, adds to markedness theory (a) the concept of *uncertainty* that low-frequency information may be predictable (e.g. all occurrences are found in one author) or unpredictable (e.g. all occurrences are found across many authors),²² and (b) that such predictability affects the markedness of the linguistic item.
- It should be understood that UC(R) despite its advance billing, does not look *only* at low-frequency information. Because low-frequency information suffers less noise losses, it necessarily has a higher information content (per SP₁₀). Thus UC(R) simply tends to weight low-frequency information more highly than does standard Fisherian metrics such Cramér’s V or G&K’s tau.²³

²¹ The theory of markedness arose from the Prague school of linguistics. Markedness assumes that language systems are organized as sets of binary oppositions. This construct is used in SFL, and has deeply influenced Porter’s theory of Greek aspect. See Porter, *Verbal Aspect in the Greek of the New Testament*, 245–51, and Porter, *Idioms of the Greek New Testament*, 23–25. While any exclusive mathematical construct can be explained using a series of binary oppositions, nothing in either linguistics or mathematics demands that language should be essentially or even primarily viewed as a set of binary (dichotomous) oppositions. By almost any measure (information storage, search speed, number of nodes, modeling parsimony and the like), binary systems are one of the most inefficient ways to build a data system.

²² This raises an intriguing question worthy of further investigation. Might genre and authorship represent two poles of a cline—what is predictable in one is unpredictable in another.

²³ In particular, because IT-based measures (unlike chi-squared measures) rely on log-linear mathematics they are based on observed rather than expected frequencies in a cell. Accordingly, they are far less influenced by tables with large numbers of cells with low expected values. While interpreting sparse tables via chi-square statistics is certainly contraindicated, this critique must be very clearly distinguished from the common misperception quite unfortunately

- We now have all the necessary definitions in place to interpret UC(R) correctly. If the UC(R) is higher, more of the total language information sent by the sender is able to be predicted by the genre or authorship theory. If the UC is lower, less of that information is able to be predicted. In our specific case in which all authorship or genre theories are being compared with reference to the identical language measure, a lower UC(R) means that more unique information has been destroyed or made equivocal. Said in Krippendorff's language, there are a larger proportion of "surprises" in the language measure that that specific grouping *can't* predict.

With this background, we now have all the resources we need to understand the findings of the sections which follow—and how information theory can provide us with, well, *more information* than standard Fisherian statistics regarding authorship and genre in the GNT.

repeated by Wellek who inveighs against *all* deviation-based statistics for their "focus on deviations from, *and distortions of*, [italics mine] the linguistic norm." René Wellek, "Closing statement (retrospects and prospects from the viewpoint of literary criticism)" in Sebeok, ed., *Style in Language*, 417-18. Halliday attempts to adequately address Wellek's concern (Halliday, *Linguistic Studies of Text and Discourse*, 99). To Halliday's response we would add that Wellek (a) seems unaware that standard statistical praxis based on classical Fisherian distribution theory actually *prevents* distortions by limiting interpretation of contingency tables with a large percentage of low-frequency cells, (b) appears unaware of information theory (IT) which provides a complementary "low-frequency" view of the distribution, (c) wrote apparently without a knowledge of the then extant field of Bayesian inference which yields a third largely independent method for inference and (d) made his conclusions, apparent without knowledge of exact tests (see Agresti, "A Survey of Exact Inference," *passim*) in which significant deviation in contingency tables can be calculated exactly, rather than being based on a distribution. Collectively, these advances comprehensively mitigate Wellek's concerns regarding the limitations of the putative distributions of Fisherian statistics. Gadamer, in contrast, fully understands how statistics may be misused, even as propaganda. Yet, he also understands that correct statistical interpretation is a hermeneutical inquiry—the right questions addressed by the proper statistical method and interpreted in light of that method's *limitations* (Gadamer, *Philosophical Hermeneutics*, 11).

9.3 Exploration One: Does Authorship or Genre Contain the Most Information?

To precisely answer this question using IT requires inspecting the relationship between entropy, “lost” and “added” information, and then comparing that to the total transmitted information. I have executed this analysis on all language measures, but for simplicity and for brevity present only the lemma data here:

Findings

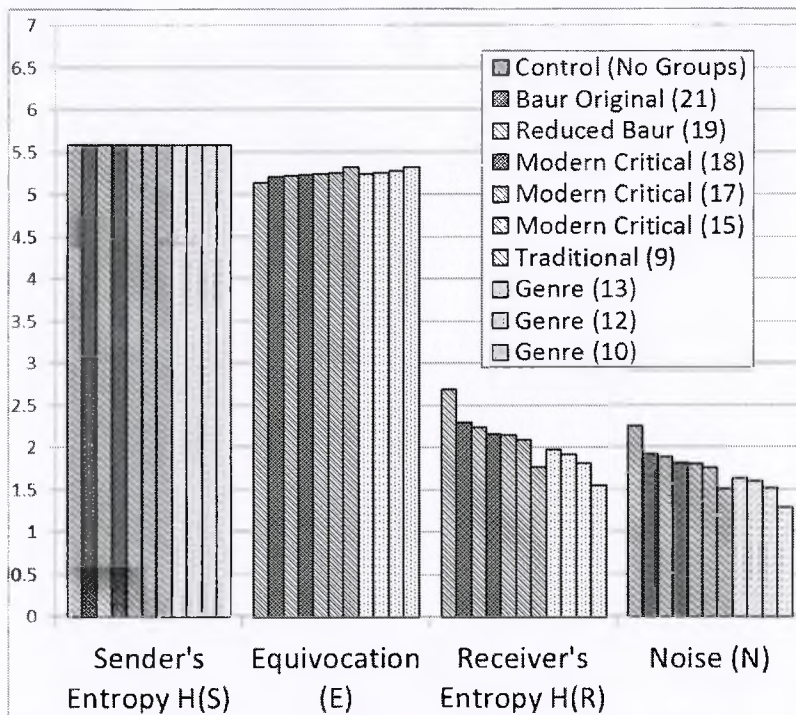


Table 9.2. Magnitude of Selected IT Components: All Lemmas of the GNT by Authorship and Genre Theories

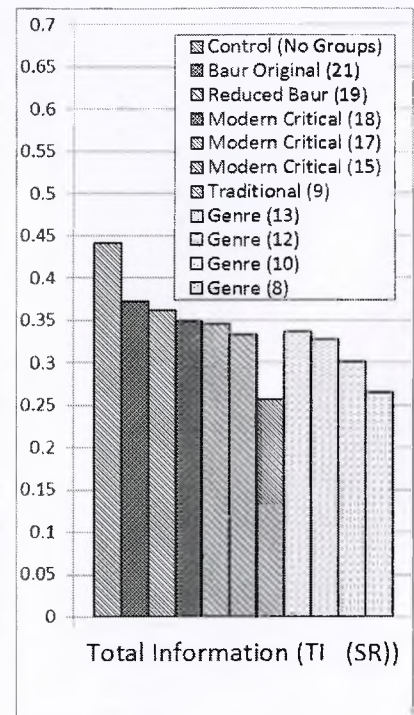


Table 9.3 TI. Lemmas by Authorship and Genre Theories

Discussion

The Sender's Entropy $H(S)$: Per SP_5 , all $H(S)$ are equivalent for all authorship or genre theories because the message being sent is exactly identical, all contingency tables have the same number of rows (5,413 categories of lemmas) and observations (the 138,019 lexemes in the GNT).

Equivocation: First recall that equivocation is the inability of the receiver to distinguish messages from the sender.²⁴ Per Table 9.2, the relative amount of equivocation differs by authorship or genre theory. Equivocation is lowest for the control case where the language measure from the 27 message streams (the texts of the GNT) are being analyzed separately (one text per category). Equivocation is generally highest, as we surmised before, for genre and authorship groupings with the least number of categories (multiple texts per category). Most critically, among all the theories, *equivocation is highest for the traditional authorship theory* (see Table 9.2). Said in terms of our concern with the traditional author theory, a high level of equivocation means that the receiver can't classify correctly the source of the message. In our case the message is the language measure being analyzed and its source is the grouping category it emerges from (say Matthew vs. Mark in terms of an authorship theory). Returning to our focus on the traditional nine-author theory, *something* is causing such a high level of equivocation. Moreover, *it isn't merely due to the fact this theory has fewer categories because (see the Section on TI below) the Genre 8 category has fewer categories but (a) a higher TI (Per Table 9.6 and especially per Table 6.7) and (b) all the genre categories have a substantially higher UC(R) than does the traditional nine-author category. See especially Table 6.7.*

Receiver's Entropy and Noise: Two observations stand out. First, both the receiver's entropy and noise are lower than the sender's entropy and equivocation, respectively. Second, noise is only about 80 to 85 percent of the receiver's entropy. Because the information content is the same for both the receiver and the sender, the uncertainty coefficient for the receiver ($UC(R)$ ²⁵) is necessarily higher than for the sender: $UC(S)$. Mathematically, it can be demonstrated that with

²⁴ See Krippendorff, *Information Theory*, 20–21.

²⁵ Caution is needed. Here, for consistency, the "R" in $UC(R)$ means receiver not row. In row/column parlance our $UC(R)$ is called $UC(C)$ for "column". What we denote as the sender uncertainty coefficient, $UC(S)$, is denoted in row/column parlance as $UC(R)$ where the "R" means row.

H(S) being constant, *equivocation*²⁶ or *noise*²⁷ in the GNT is due to the different number and nature of the columns created by the theories, $\Delta H(R)$, plus the interaction of the authorship or genre theories with the language measure: $\Delta H(SR)$. The operative phrase here is the last one—while equivocation is certainly increased by a reduced number of categories it is also affected by the differential interaction of authorship and genre with the language measure.

Total information: Interpretively, the TI:(SR) values in Table 9.3 are very revealing. Since TI:(SR) can only change due to a change in H(R) and/or H(SR), per SP₇, the simple act of combining texts together either directly or indirectly destroys information. Said equivalently, combining texts increases noise and hence destroys information. Accordingly, it is to be expected that traditional authorship and the eight-category genre theory display the highest amount of destroyed information and hence the lowest overall amount of information within their class of theories. *But the level of the destruction of information as evidenced by the higher equivocation in the traditional nine-authorship theory is more than would be expected by this grouping phenomena alone.*

Implications

We actually have found a smoking gun. It is equivocation—and equivocation is differentially highest for the nine-author theory. This strongly supports the line of reasoning in 7.4.

²⁶ Given that $E=H(S) - TI:(SR)$, per GP₆, and given that H(S) is constant, the change in E (ΔE) can only arise from $\Delta TI:(SR)$. Moreover, given SP₆, ΔE can arise only from some combination of $\Delta H(R)$ and $\Delta H(SR)$.

²⁷ Given that $N=H(R) - TI:(SR)$, per GP₈, the change in N (ΔN) can only arise from $\Delta TI:(SR)$. Moreover, given SP₈, ΔN can arise only from some combination of $\Delta H(R)$ and $\Delta H(SR)$.

9.4 Exploration Two: Which Language Fraction Encodes the Most Information?

We address this question in tables 9.4 and 9.5 which plot TI:(SR) from four instance partitions from two separate language measures. Table 9.4 reflects TI:(SR) from semantic subdomains and Table 9.8 reflects TI:(SR) from standard clause complexes.

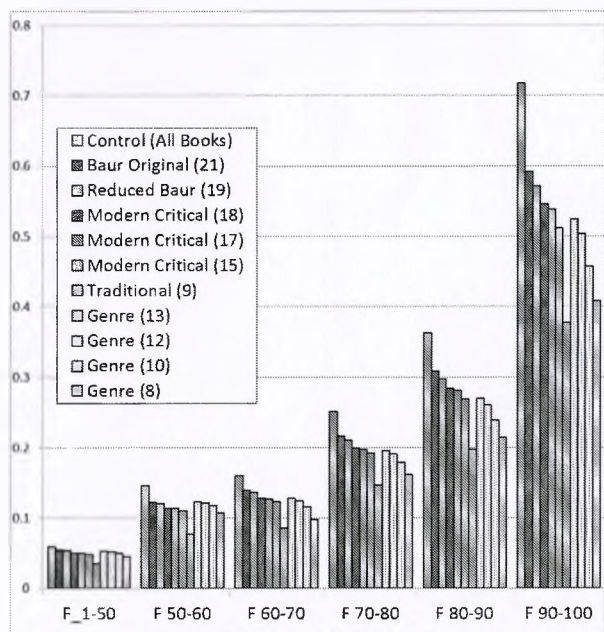


Table 9.4: TI of Semantic Subdomains by Theory by Six Frequency Deciles

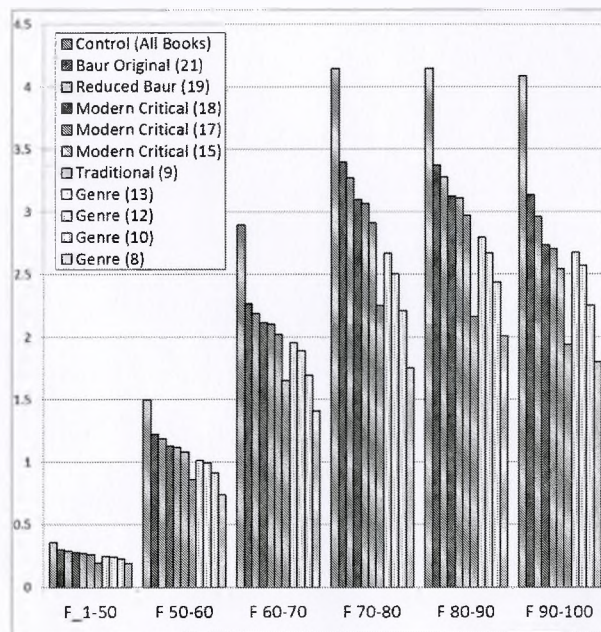


Table 9.5 TI of Clause Complexes (Standard Attribution) by Six Frequency Deciles

Table 9.4 and Table 9.5 exemplify the general case across all ten language measures.

Specifically, with lower frequency data partitions both $(H(S) + H(R))$ increase, noise (N) drops to zero and the total information, TI:(SR), expressed as a percent of the total entropy, $(H(S) + H(R))$, increases. The quicker rate at which clause complexes reach maximum total information is simply a reflection of the fact that the lowest 33.58 percent of the clause complexes formed by frequency are *hapax legomena*, compared to the lowest 0.3623% for semantic subdomains.²⁸

This underscores the axiom that unique information communicates maximal information content.

²⁸ Said in terms of category count rather than cumulative frequency, only 0.603% of the semantic subdomain categories are *hapax legomena*, but 93.79% of the clause complex categories are *hapax legomena*.

Implications for Genre and Authorship

Why should we care if more information is transmitted by the less frequent partitions of language? For one simple reason. Low-frequency information (which may be idiolectic/idiosyncratic but tend to be more generic/context specific) is a manifestly underutilized resource in allowing us to predict “markers” that are markers in classification tasks. I trust by now that this will not be seen to be contradictory but rather complementary to our prior statements regarding the higher overall utility of high-frequency species in authorship attribution tasks. Let me restate this point clearly. High-frequency patterns (e.g. the patterns of the fifty most frequent marker words) are disproportionately useful for genre or authorship classification *as well as* unique low-frequency forms. The apparent antinomy can be resolved as follows: The former are the preferred fodder for MLT classification techniques and the latter by infrequently-used IT techniques. In future stages, through our abductive cycle, multivariate IT techniques that are disproportionately sensitive to low-frequency fraction language are clearly indicated as we seek to sharpen the boundaries between genre and authorship in the GNT.

9.5 Exploration Three: How Does TI Loss Differ by Language Measure in the GNT?

Findings

One may ask, per the prior section, why should we care if TI loss in the GNT differs by language measure? The primary reason for doing so is to understand if the differential loss of information by the various authorship or genre theories derives disproportionately from one “part” of the linguistic landscape or the other. To discover this requires that we somehow normalize the TI in each partition, since, as can be seen from Tables 9.6 and 9.7, TI:(SR) increases as language information decreases. Given that all genre and authorship theories are different text groupings of the control case, a legitimate and obvious approach is to normalize all partitions by dividing the TI of each authorship or genre theory by its own control. This expresses each theory as a percentage of the total information of the “control case” in each partition. Doing so for two of the linguistic measures yields the results in Table 9.6 and Table 9.7.²⁹

²⁹ See TI_10Ling_x_CatQuart_x_Theories tab in the spreadsheet: ftp://decisionsupportsciences/stat_ling/Ling_by_FourQuartiles_by_Theories.xlsx.

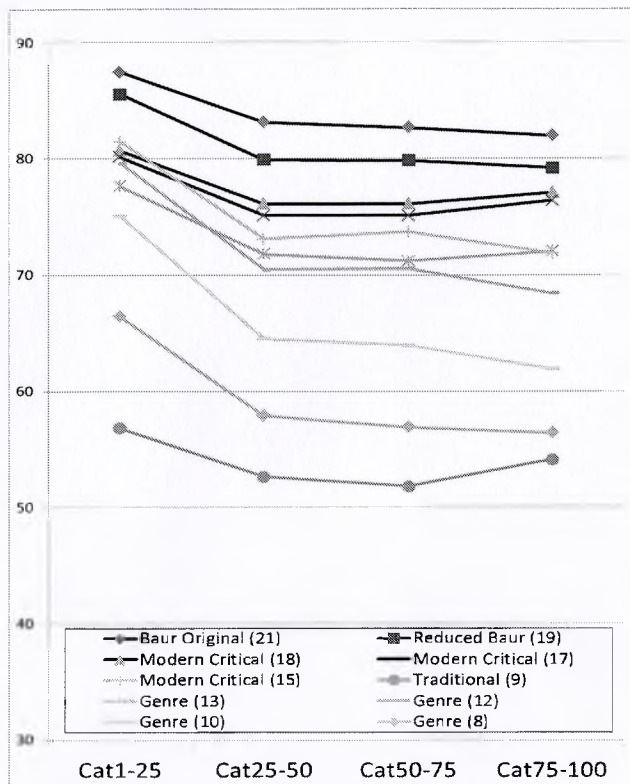


Table 9.6: Total Information, TI:(SR), of Semantic Subdomain Category Quartiles Expressed as a Percentage of Their Control Group (All Texts)

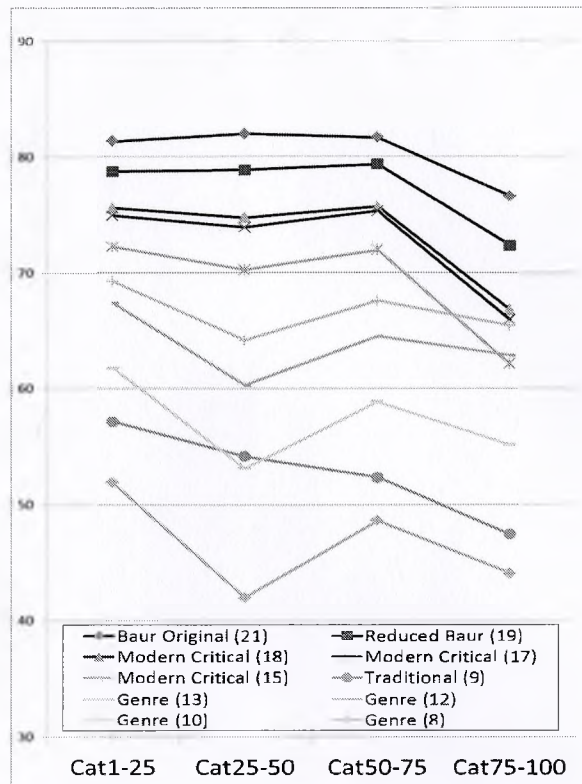


Table 9.7: Total Information, TI:(SR), of Clause Complex Category Quartiles (Standard Attribution) Expressed as a Percent of Their Control Group (All Texts)

Discussion of Findings:

Table 9.6 demonstrates that when the semantic subdomain quartiles are expressed as a percentage of the control category (the control is where each book is its own category) the traditional theory of authorship displays the *lowest* amount of total information in every quartile compared to all other theories. Table 9.7, by contrast, demonstrates that when standard attribution *clause complex* quartiles are normalized, Genre 8 displays the lowest amount of total information in every quartile compared to all other theories. Visual inspection of both tables, moreover, confirms that for clause complexes the four genre theories encode significantly lower total information than with semantic domains. That is, as a percentage of the total information communicated, semantic domains communicate a higher percentage of genre information and

clause complexes communicate a higher percentage of authorship information. It is clear, then, that different linguistic measures *as a class* seem to function differently with respect to authorship and genre. This in turn provides strong initial empirical support for the broad functional linguistics contention that *different levels of rank encode different types of information*. Or, alternatively, could this difference be because semantic information operates at a different stratum (semantics and not lexicogrammar)?

One more observation is relevant. Recall that (per Figures 7.4 and 7.5) high-frequency clause complexes are the “best place to look” for authorship markers. That is, 16 out of the 20 HLLA models in which authorship was stronger than genre occurred at the rank of the clause or above.³⁰ Notice moreover, that in Table 9.7 all the genre theories lose more information (are down-shifted) compared to lemmas (Table 9.6). The differential loss of genre information in clause complexes compared to lemmas may simply be because structures at this rank “carry” a higher proportion of idiolectic information over sociolectic information. Said another way, clauses and clause complexes may be more culturally suited in the GNT for carrying idiolectic information. In this case the TI for authorship theories would not experience as much down-shifting as genre. If it is indeed true that we have discovered that different linguistic measures carry more idiolectic as opposed to sociolectic information, this is quite important given that one of the central tasks of functional linguistics is to decode and describe how different linguistic measures function in texts. Because of the potential importance of this finding I will promote it to an exploration in cycle two of the abductive design.

³⁰ Recall that there were 54 models in this experiment, so even in authorship’s strongest language fraction, genre was still stronger: 34 models to 20.

9.6 Conclusions: Initial Information Theory Explorations into Authorship and Genre

In this chapter the goal was to search for a smoking gun as to why the traditional authorship theory differentially lost the most information compared to the other authorship theories. That smoking gun is equivocation. Not only was it discovered that equivocation is indeed maximal in the case of the traditional authorship theory (9.3), the data demonstrated that equivocation or noise *is due to the different number and nature of the columns created by the theories, $\Delta H(R)$, plus the interaction of the authorship or genre theories with the language measure: $\Delta H(SR)$* . This is the interaction that we can tease apart using techniques such as HLLA analysis, but the larger point is that *$\Delta H(SR)$ varies depending upon the theory*. This is what allows equivocation to be higher in the nine-group traditional authorship theory, when in fact we surmise that it should be higher in the Genre 8 theory (see Table 9.2).

More information is communicated by the low-frequency GNT partitions, and combining texts into fewer categories causes equivocation which differentially destroys low-frequency information (per Tables 9.5 and 9.5).

The total information communicated by the various theories changes when looking at different language measures. This clearly seems to point to the fact that language measures of different rank differentially encode different amounts of information regarding authorship and genre (9.5). If it is indeed true that we have discovered that different linguistic measures carry more idiolectic as opposed to sociolectic information, this is quite important given that one of the central tasks of functional linguistics is to decode and describe how different linguistic measures function in texts.

10 Summary Conclusions and Hermeneutical Reflections

Certainty is, *rebus sic stantibus*, unattainable.

D. Wickmann¹

High probability leading to psychological certitude is *not* unattainable.

Gordon R. Lewis²

In this work I explored linguistic style in the GNT and tested the extent to which its presumed causes, specifically authorship and genre, can be “disentangled.” To do so a research plan was implemented in six steps. *First*, a linguistically structured database of the GNT was developed—one that annotated the text of the GNT paradigmatically (by system), syntagmatically (by structure) and semantically (by strata).³ *Second*, various historical theories of authorship and notional theories of genre were assembled in order to test them head to head. *Third*, the first analysis probed into the bivariate association, direction and information content of each of these theories for each of the language measures just described. *Fourth*, genre variation was disentangled from authorship variation using nonparametric multivariate analysis and then each authorship theory was retested against each genre theory to see whether genre or authorship explained more of the total variation in the GNT. *Fifth*, the 27 texts (books) of the GNT were clustered lexically, paradigmatically and stratally using a form of extractive multivariate analysis (Correspondence Analysis). *Next*, each theory of authorship and genre was tested to see how well it fit the actual locations of the GNT texts defined by that Correspondence Analysis. *Sixth*, given these answers I sought to think through the impact of this kind of work on NT interpretation.

¹ Wickmann, “On Disputed Authorship,” 34.

² Per Prof. Gordon Lewis, in his Systematic Theology course, Denver Seminary, 1986.

³ That is, paradigmatic, syntagmatic and stratal annotations are used in this work.

The task in this final chapter, then, is quite conventional. First, I will draw together the findings from the research. Second, implications for NT studies will be considered. Third, consonant with the subtitle of this work I will speak to method. Fourth, I will propose next steps. Fifth, I will conclude with a hermeneutical reflection.

10.1 Findings Summary

While the central objective of this work was to disentangle authorship from genre, by the time our work was done, experimental data was developed in five separate areas of NT stylistics: (1) the utility of *hapax legomena*, (2) the relative strength of genre vs. authorship in the GNT, (3) a rank ordering of the authorship theories by the amount of variance they explain in the GNT (and by extension which appears the most primordial), (4) a similar rank ordering of which genre theory accounted for the most stylistic variance in the GNT and (5) early evidence that computational stylistics can contribute to NT discourse analysis.

Each will be looked at in that sequence.

10.1.1 Finding One: *Hapax Legomena*

These findings may come as a surprise for those not familiar with recent computational stylistics research regarding *hapax legomena* in mixed-genre texts. To frame the research findings in this work, first recall the summary findings presented in the review:⁴

- 1) *Hapax legomena* have fallen from favor in terms of their utility in authorship studies.
- 2) *Hapax legomena* and low-frequency words in general are increasingly being understood to be more reflective of specialized registers required by specialized contexts. The prior monolithic view that *hapax legomena* is a stolid marker of authorship is to be rejected.
- 3) High-frequency words are now definitively accepted as better markers of authorship than genre. Specifically, the two most successful authorship approaches in the subsequent history of research into authorship attribution have been:
 - a. The high-frequency “function words” of the “Burrows’ method” (late 1980s to the present)

⁴ See section 3.3.2.3.

- b. The high-frequency, largely lexical “feature-vectors” used by the current generation of authorship attribution methods (mid-1990s to the present)

In addition to results developed by others, this study discovered that *hapaxes* have their place—but not in terms of authorship. Consistent with Greenspahn this work discovered that low-frequency words are substantially more associated with genre than authorship.⁵ Two factors give rise to these *hapaxes*. First, novel vocabulary is the natural by-product of a distinct context because it calls forth a different register. Second, and not so obviously, when these single texts are combined into a corpus, register-specific language (e.g. vocabulary) that is decidedly not low-frequency within a homotypic corpus will *become* lower (relative) frequency language within a heterotypic corpus.

NT scholars will immediately ask what impact our findings have upon the 1921 findings of P.N. Harrison. Along the lines of Herdan’s critique,⁶ and given the lines of research defined above, even if P.N. Harrison’s *hapaxes* were collected correctly (and they were not), counted correctly (and they were not), and even if they were normed for the size of the text (and this would await Zipf’s work), his *hapaxes* are very likely populated with genre-associated novel vocabulary. While a definitive retrospective computational accounting of Harrison’s work still awaits, given our findings, Harrison’s work, assiduous as it may have been, almost certainly disproportionately mistook register-specific *hapaxes* to be idiolectic in nature.

10.1.2 Finding Two: More of the Total Stylistic Variation in the GNT is Explained by Genre Than Authorship

This, the epicenter of our work, confirmed that not only is stylistic variation due to genre present in the NT, but that *more of the total summed stylistic variation is due to genre than authorship*. This finding holds regardless of whether that variation is (1) assessed *in toto* by

⁵ See section 7.3

⁶ Grayston and Herdan, “Authorship of the Pastorals,” 4–7.

nonparametric measures of association (such as Cramér's V, G&K tau or the Uncertainty Coefficient), (2) decomposed into its main effects and interactions and parametrically compared using hierarchical log-linear analysis (HLLA), (3) decomposed into components using EMVA (CA) for the purpose of qualitative visualization or (4) the individual authorship or genre theories tested for goodness of fit to the locations of the NT texts in the resulting GNT space. The ubiquity of these results is striking. Whether by higher association in the bivariate analysis, greater reduced chi-square in the HLLA results, the qualitative assessment of the CA reduced spaces or the lowest Euclidean distance spread, genre is predominant over authorship. This leaves us little alternative but to reassess prior stylistic studies in pseudepigraphy and authorship. Specifically—and this is the major finding of this study overall—to the extent that prior work has either (1) failed to test for genre as a competing theory or (2) failed to remove the effect of genre as a covariate, those studies have almost certainly confounded genre with authorship.

10.1.3 Finding Three: The Traditional Authorship Theory Explains More of the Total Variance in the GNT than Any Other Authorship Theory

As a consequence of finding four, non-parametric models (HLLA) were built which explored the interaction between the language of the NT (via the nine language measures) and genre and authorship. This analysis itself tested each of the six authorship theories with each of the four genre theories. This generated two interaction terms of interest: the interaction of the language of the NT with authorship disentangled from genre (A|G) and the interaction of the language of the NT with genre disentangled from authorship (G|A). In effect this approach uses mathematical surgery to separate authorship from genre. *Upon routine inspection of the normed statistics it was discovered that the authorship theory that accounts for the most information with the strongest genre theories was always the traditional authorship (the nine-author theory).* This did not occur occasionally but in all 108 models that were tested head-to-head against the

strongest genre theories. What makes this even more disruptive, at least in terms of the current mainstream authorship theories in NT studies, is the implication of this finding. All else being equal, the strongest authorship theory disentangled from genre effects has the highest likelihood of being the primordial theory of authorship. *The difficulty of the colligative strength of these findings is that it places us in an interpretive bubble of sorts.* On the one hand, I do not wish to downplay the strength of these findings regarding authorship. On the other hand the research has only completed its inaugural cycle of abduction. How, then, do we express the strength of these conclusions with the experimental caution that *must be* the hallmark of post-positivist interpretation? It is here where the rationale for choosing IBE comes to the rescue. After the completion of *each* analysis cycle IBE expects us to actually make an inference to the best explanation. Accordingly, *given what is known now, (and recognizing that no information will ever be complete!) what is the inference to the best explanation?* Accordingly, I repeat the analogy made in Chapter Seven—if the issue of New Testament authorship can be compared to a horse race with our sequential experimental design as the racetrack, we are indeed in the early stages of that race—but the most ancient horse just took the lead.

Before moving on, one more facet of this finding should be considered. If this authorship finding becomes more confirmed, a rich interpretive framework for understanding this data emerges with the combined aid of information theory (IT) and SFL theory. First, it was discovered via information theoretic explorations in Chapter Nine that although the traditional authorship theory is the strongest theory of authorship as measured by Cramer's V, G&K tau and HLLA, *it is also the single theory with the lowest UC (an IT measure).* Why? Information theory plus SFL provides an explanation as follows.

First let's look at "mixing." According to the nine-author theory, one single author in the NT, Paul, wrote across six of the thirteen genres in the GNT, Luke wrote across two genres, John across three genres, and Peter across two genres. *This mixing of genres is far greater than in any other authorship theory.* Because of the substantial interaction between authorship and genre, this is simply a recipe for what information theory terms equivocation—and on a massive scale. Now let's look at what SFL can contribute. Equivocation makes it hard (actually impossible) for IT to predict from which category the original "message" came. The texts of the NT, by all accounts, arose occasionally to meet the needs of the theological communities they would serve. Dozens and perhaps hundreds of separate contexts required almost as many specialized registers to be employed by the authors/editors/redactors of these texts. These registers which occur in high frequency in their own setting became decidedly low-frequency registers when combined into the corpus of the New Testament. When the data is mixed together (as it is in its native habitat within the NT), but then grouped by the traditional nine-group authorship theory and cross tabulated against a given language measure (without disentangling it from genre), then *the authorship theory where the authors write across the most genres will naturally be the one that is most difficult to predict by information theory.* As a confirmation of this hypothesis the research explored if the preponderance of low-frequency information (the fraction that information theory "weights" the most) was genre. The HLLA analyses confirmed this finding.

10.1.4 Finding Four: Genre 13 Explains More of the Total Variance in the GNT than Any Other Genre Theory

The argument for genre is similar to the above, but I wish to point out three facets of the genre data. First, the bivariate "entangled" distribution tells us a modestly different tale than the disentangled data. What is not in doubt is that all theories of genre explain more of the language variation in the GNT than all theories of authorship, and genres with fewer categories (e.g. Genre

8) are “stronger” via bivariate association.⁷ *When genre is disentangled from authorship, however, the strongest genre categories in terms of authorship are Genre 13 and Genre 12.*⁸ This may imply that the higher association seen with Genre 8 in the “entangled” bivariate data may be because it is strongly covarying with authorship. More research here is indicated. Second, our original development of four genres followed a type/audience/number typology drawn in essentially equal parts from literary theory, genre criticism, and rhetorical criticism.⁹ The audience dimension and the number dimension, however, were motivated by SFL’s notion of the clause as an interpersonal exchange between a speaker/writer and an addressee.¹⁰ Whether we call it audience (per Bell) or addressee (per Halliday) is unimportant.¹¹ What *is* important is that an interpersonal exchange rich with linguistic implications is implied, including the form of address and the mood of predications (the level of finiteness) chosen.¹² These in turn are strongly governed by the *status* of the audience addressed or the status of the speaker/writer in relation to that audience.¹³ My point is that although I have tried to build these kind of ideas into the definition of genre used here, our “type/number/typology construct” or even Pitts’ similar notion¹⁴ hardly exhausts the refinements that can be made to this modest starting set of four genre configurations. Third, and related to the point just mentioned, late in the research cycle two new genre configurations were added to the mix suggested by the conformation of the clusters themselves: Genre 12E and Genre 13E.¹⁵ These two empirically derived categories of genre

⁷ See Tables 6.1-6.6.

⁸ See Table 7.2.

⁹ See Section 5.4.

¹⁰ Halliday and Matthiessen, *An Introduction to Functional Grammar*, 106–11.

¹¹ See Bell, “Language Style as Audience Design,” 158–72; Halliday and Matthiessen, *An Introduction to Functional Grammar*, 135.

¹² Halliday and Matthiessen, *An Introduction to Functional Grammar*.

¹³ See Bell, “Language Style as Audience Design,” 150–54.

¹⁴ Pitts, “Style and Pseudonymity in Pauline Scholarship,” 133–38.

¹⁵ See Figure 8.3.

substantially outperformed other genre categories lexically and semantically, but not at the rank of the clause and above.¹⁶ This kind of rich interoperation between theory-driven and empirically-driven category improvement is welcomed and expected in abductive designs.

10.1.5 Finding Five: Computational Stylistics Can Contribute to NT Discourse Analysis

While the data is thin here, it *is* directional. In the process of attempting to discover the functional importance of clause complex *hapaxes*,¹⁷ I discovered that the books in the GNT which had the *highest frequency* of these *hapaxes* (a la Harrison's practice)¹⁸ were Romans, 1 Corinthians, 2 Corinthians, 2 Thessalonians, James, 1 Peter, and 2 Peter. These all derive from one genre, *epistles*, and more specifically from a subgenre, *corporate epistolary*.¹⁹ Further inspection reveals that these *clause complex hapaxes* were infrequent for a singular reason: their clause chains were very long. From a discourse perspective, long chains of clauses typically achieve paraenetic²⁰ or persuasive goals in discourse—which of course, is one of the signal functions of corporate epistolary genre in the GNT. What was discovered in fact was that functionally similar sections of NT discourse sections emerged from computationally similar clausal complex *hapaxes*. When analyzed, what was discovered was quite intuitive. This seems to be appraisal language which gives rise to structured arguments, which in turn use long syntagms (long spans of linguistic units) as one of their central devices. Thus, computational stylistics, precisely because it “sees” things we do not see—may be used as a new resource in NT discourse analysis.

¹⁶ See Figure 8.7. More research is indicated to discover why the outperformance of Genre 13E and Genre 12E lessened with increasing rank.

¹⁷ See Section 7.4.

¹⁸ These had a frequency 1.2 times greater than the average frequency in the GNT.

¹⁹ The *individual* epistles, specifically the Pastoral Epistles, displayed significantly *fewer* clausal *hapaxes*, at an average rate of 0.451, 0.364, and 0.488 for 1 Timothy, 2 Timothy and Titus respectively.

²⁰ Unlike Stowers, I use paraenesis here to embrace both its protreptic (to encourage toward a new way of life) and its paraenetic senses (to encourage toward continuing the current way of life). This is closest to its ancient usage. Cf. Stowers, *Letter Writing in Greco-Roman Antiquity*, 92.

10.2 Implications for NT Studies

10.2.1 Implications of the Predominance of Genre Over Authorship in Terms of the Pauline Canon

As discussed in the second finding of this chapter, genre accounts for more of the total variation in the NT texts than authorship across every bivariate and multivariate method executed in this work.²¹ Beyond what the data *say*, it is time now to delve more deeply into what it *means* for NT studies. Several points must be stated as clearly as possible so that we can feel the full weight of these findings. First, these findings demonstrate that the long-standing *de facto* assumption that virtually all stylistic variation is authorial in source *is simply false*. Second, because genre variation is actually *greater* than authorship variation, whenever scholars have asserted that the stylistic differences are “too great to be caused by authorship” they now *must* add, “but likely not too great to be explained by genre.” Scholars who have concluded against, for instance, Pauline authorship on the grounds of novel vocabulary or perceived non-Pauline content, have only three options to attempt to resuscitate the line of reasoning that a different style means a different author. First, they can control for genre mathematically (as we have done).²² Second, they can perform supervised multivariate analysis of the larger Koine to collect lexical, paradigmatic or semantic forms that do separate authors well and then look for that profile in the GNT. Third, they can forget the math and control for genre by eliminating it entirely by analyzing only homotypic texts as Burrows did. Problematically, while the last strategy is possible in larger corpora, no such convenient doppelgangers exist in the NT. As it

²¹ In Binongo’s words, it predominates over authorship: “When the essays and plays are brought together into one picture...the difference in genre predominates over other factors.” Binongo and Smith, “A Bridge between Statistics and Literature,” 784.

²² This will allow for the extraction of genre- or authorship-specific language items or fractions of the same.

has been conventionally argued,²³ then, scholars who have used large style variations to argue for pseudepigraphy or pseudonymity have likely been building another structure altogether—one that wholly subverts their intentions. Some representatives of this line of argumentation include its inaugurator J.E.C. Schmidt (1 Timothy),²⁴ Baur (all Paulines but the Hauptbriefe),²⁵ Goodspeed (Ephesians),²⁶ and Harrison (Pastorals).²⁷

Lastly, the CA data demonstrates persistent cluster groupings that make it more difficult to hold unwaveringly to the mainstream authorship perspective on the Pauline Canon. Specifically, the closest text pair that most closely and regularly pairs together across all texts in the GNT and across all language measures is the pairing between 1 and 2 Thessalonians. Subsequent research (not shown here) demonstrates this pairing is persistent as well when the analysis focuses only on the twenty-one texts of the GNT that may (in the broadest sense) be classed as epistles. While we have not yet seen a clear authorship component extracted from this data—the search is still ongoing—the very minimal distance between 1 and 2 Thessalonians clearly problematizes the assignment of 1 and 2 Thessalonians to different authors.

10.2.2 Implications of the Predominance of Genre Over Authorship in Terms of the Pseudepigraphy and the NT

The usefulness of computational stylistics in speaking to contemporary issues in NT pseudepigraphy is more nuanced and in a certain sense more difficult to assess. Two reasons make this so. *First*, much of the modern era relevant to pseudepigraphy, if we date it somewhat arbitrarily from Holtzmann, has focused on the *intention* of the pseudepigraphist and the

²³ One argument is viable, to demonstrate that the total variation is greater than expected by the maximal value of authorship variation *plus* the natural sum of genre variation. This too requires decompositional methods, and very few pairs or groups of texts are this different.

²⁴ Schmidt doubted both epistles on the basis of aberrant vocabulary, and Schleiermacher because he viewed it as unknown to Polycarp as well as to the apostolic fathers; Schmidt, *Historisch-Kritische Einleitung*, 257–67).

²⁵ Dibelius, *Die Pastoralbriefe*, 3

²⁶ Goodspeed, *The Key to Ephesians*, vi.

²⁷ Harrison, “Important Hypotheses Reconsidered,” 77–81; Harrison, *The Problem of the Pastoral Epistles*.

reception of his or her product. Opinions on the former span the spectrum as follows: real forgeries (Beatrice), deceitful practices (Duff), considered attempts to deceive (Baum), noble lies (Brox and Donelson), pious frauds (Candlish), transparent literary fictions (Frey), something less than forgeries (Jülicher), nondeceptive works forged in theological crisis (Riedl, Sint), and works of modesty or unstained piety (Moffatt).²⁸ Of what possible benefit can computational stylistics be, one might ask, in shedding light on the *intentions* of the long dead pseudepigraphist? Surprisingly, however, recent supervised multivariate analysis has, in fact, rapidly improved the ability of researchers to detect plagiarism,²⁹ and SFL, in my view, has emerging assets of tenor that may indeed be capable of discovering style shift indicative of such an intent.³⁰ That said, this remains only an intriguing potential future direction possible for NT computational stylistics.

Second, pseudepigraphy is a conclusion not a cause. That said, it is *based* on causes, or at least rationales, and often quite a number of them, with style simply *one* of those causes. If we are to attempt to simplify the variegated landscape of pseudepigraphy, however, it seems that most scholars that conclude for pseudepigraphy draw from what might be classed as two major wells of rationales. One well is more related to synchronic concerns: namely stylistic, literary, linguistic and grammatical factors. The second well is more related to quasi-diachronic concerns such as source critical, doctrinal, theological, ecclesiastical and even text-critical factors.

²⁸ Beatrice, "Forgery, Propaganda and Power in Christian Antiquity," 49; Duff, "A Reconsideration of Pseudepigraphy," 306–09; Speyer, *Die literarische Fälschung*, 171–303; Baum, "Literarische Echtheit als Kanonkriterium," 97–110; Brox, *Falsche Verfasserangaben*, 81–105; Donelson, *Pseudepigraphy and Ethical Argument*, 9–23; Candlish, "On the Moral Character of Pseudonymous Books: II," 272; Frey, ed., *Pseudepigraphie und Verfasserfiktion*, 489–536; Riedl, *Anamnese und Apostolizität: der Zweite Petrusbrief*, 238–39; Moffatt, *An Introduction to the Literature of the New Testament*, 41. Baum, Beatrice, Duff, Bronx, Frey, Riedl and Sint are as cited in Baum, "Authorship and Pseudepigraphy," 56–63.

²⁹ See especially the recent work of Kuta and Kitowski, "Optimisation of Character N-Gram Profiles," 500–11; Oberreuter and Velásquez, "Text Mining Applied to Plagiarism Detection," 3758–63; Tschuggnall and Specht, "Using Grammar-Profiles to Intrinsically Expose Plagiarism in Text Documents," 298–302; Stein et al., "Intrinsic Plagiarism Analysis," 63–82 and Stamatatos, "Plagiarism Detection," 2515–26.

³⁰ See Davis et al., "Stance Analysis of African Americans with Diabetes," passim; Davis and Mason, *Locating Presence and Positions in Online Focus Group Text with Stance-Shift Analysis*; Rickford and McNair-Knox, "Addressee- and Topic-Influenced Style Shift," passim; Bell, "Language Style as Audience Design," passim.

Complicating things further, scholars quite often do not provide a full accounting of their rationale at all. Given these complications at least this much can be said: To the extent to which stylistic grounds have been used as the predicate for pseudepigraphal conclusions, such arguments for pseudepigraphy now appear to be weakened. Several examples will suffice to make the point—but the story is a mixed one because scholars are quite varied in their relative dependence upon the linkage between style and pseudepigraphy. To begin with, the rationale of Evanson and de Wette, who viewed Ephesians as a copy of Colossians,³¹ cannot be directly impugned by our findings. Pastiche, the practice of intentional copying, is indeed detectable using modern computational stylistics methods but that is another set of analyses entirely. Baur's argument, on the other hand³² (for the few that still hold to it) has been weakened since it was, in part, based on phrasing deemed pseudepigraphal.³³ Jülicher, though cautious in terms of pseudepigraphy, nonetheless concluded that the style of the Pastorals evidenced a patchwork product.³⁴ His assessment, though, was based on a perceived semantic shift—that the semantic freight of certain words in the Pastorals differed from normal Pauline usage.³⁵ Assuming such a shift is granted, the present study locates the preponderance of stylistic variation in register, and register theory does indeed embrace such shifts over time.³⁶ That is, not only can our lexical and linguistic stock evolve as time passes, but different contexts can induce us to shift our

³¹ Baur (1845) followed Evanson (1792) and De Wette (1843) in weighing against the authenticity of Ephesians due to its perceived dependence on Colossians. Baur, *Paulus, der Apostel*, 417–18; Evanson, *The Dissonance of the Four Generally Received Evangelists*, 312–13; de Wette, *Kurze Erklärung der Briefe* 79.

³² Baur's argument was broad-based and eventually dismissive, "What gives these Epistles their claim to the name of the apostle is simply... that they profess to be Pauline." Cited in Guthrie, "The Development of the Idea of Canonical Pseudepigrapha," 45–47.

³³ Baur included phrasing in his critique, although it was a minor point. Cited in Jülicher, *An Introduction to the New Testament*, 62.

³⁴ Jülicher, *An Introduction to the New Testament*, 62, 179.

³⁵ Jülicher, *An Introduction to the New Testament*, 179.

³⁶ See, in particular, Pitts' diachronic axis in his register-based profile of the Pauline Canon (Pitts, "Style and Pseudonymity in Pauline Scholarship," 133–38).

semantic center of gravity for certain words that are both polysemous and frequent in our lexical repositories. By this logic, Jülicher's rationale, then, has found, at least an opposing argument, though the data in this study was not designed to address stylochronometry. I could continue scholar by scholar, but the point, I trust, has been made. Scholars who have decided for pseudepigraphy for stylistic reasons, but have not accounted for the extent to which specialized contexts draw from specialized linguistic repositories, have likely overstated their arguments.

10.2.3 Implications of the Relative Strength of the Six Authorship Theories Tested in the GNT

The Baur-Derived Hypotheses: Inspection of the data tells a completely consistent story in terms of the Baur-derived hypotheses (Authorship 19 and 21). Whether “entangled” with authorship (Tables 6.1-6.6), disentangled (Tables 7.6), or extracted (Table 8.7) the Baur-derived hypotheses were the weakest authorship categories.

The Mainstream Authorship Theories: The two mainstream authorship theories (Authorship 17 and 18) uniformly “land” intermediate between the dual Baur hypotheses and the Traditional theory of authorship.

The Traditional Authorship Theory: The traditional authorship hypothesis, in contrast, was almost uniformly the strongest. The only exception was in the data in figure 8.7, where the hybrid Authorship 15 category slightly outperforms the traditional theory of authorship lexically and semantically, but not at or above the rank of the clause.³⁷

10.2.4 Implications of the EMVA Results on Synoptic Studies

The data is lean here because this was not a focus of my work *per se*, but two items do merit attention. First, the EMVA results, especially in Table 8.1, underscore the extent to which

³⁷ Recall that this category was inspired by Kümmel who embraced the authenticity of Colossians on the basis of style. See Table 5.3.

these results are predominated by genre. Luke, for instance, is almost five times closer to Matthew than to Acts—despite the fact that he is almost universally considered as the author of Luke-Acts. Second, and this is more speculative, intriguing possibilities exist for tagging pericopae within Matthew and Luke with Q, M, L, and Mk material. Once tagged, these putative sources can be separated using EMVA analysis or HLLA. Such analyses may allow researchers to test how successfully these putative units of the tradition entangle or disentangle. These steps, however, must await future design cycles.³⁸

10.3 Implications in Terms of Methodology

10.3.1 Method Implication One: Embracing Both the Unsupervised and Supervised Revolutions

The review in Chapter Three discovered that a veritable revolution, three of them in fact, rolled through the discipline of computational stylistics in the last thirty years: supervised multivariate analysis, unsupervised multivariate analysis and feature set selection (FSE).³⁹ No longer is it adequate to select markers *ad hoc*, test them with simple univariate methods, assign results *de facto* to authorship and defend them by referencing astronomically large values reported by simple tests of inference.⁴⁰ This practice has (mercifully!) been replaced with (1) FSE methods which discover best markers rather than propound them, (2) multiple forms of multivariate analysis that recognize that in the real world complex effects have multiple causes, (3) more conservative approaches to validation including new types of statistics, (4) new experimental designs to keep researchers honest and (5) new (even post-positivist!) approaches

³⁸ Such an activity raises the question as to the possibility of something like a synchronic *traditiongeschichte*; a synchronic quantitative analysis of the texts that can get us back to the tradition itself or perhaps even to the contexts *behind* that tradition.

³⁹ See especially Section 2.4.

⁴⁰ Even later in his career, we still find Morton writing, "...on the evidence of this one Table, the probability that Peele wrote a play of Shakespeare is less than one in forty million." Morton, "Once. A Test of Authorship," 8. Compare this practice with the unease Mosteller has in the kinds of overstatements such numbers invite (Mosteller, "The Writing Styles of the Authors of 'the Federalist' Papers," 139).

to validation. In the face of all this, the discipline of NT studies lags seriously behind. We have much, much work to do. In the face of this state of affairs Rudman⁴¹ and Argamon⁴² have called for a reworking of theory. Mealand has called for better controls.⁴³ I submit here that the dual implementation of abduction and experimental design theory are a way to integrate both.⁴⁴

10.3.2 Method Implication Two: Adopting Modern Theories of Language and Linguistic Analysis

It is quite noteworthy that when the various linguistic measures studied herein underwent extraction by Correspondence Analysis clear and fundamental differences were observed between each of the nine measures submitted and every other measure.⁴⁵ While the maximal conformational differences observed were between measures of higher rank (clauses and clause complexes) and measures of lower rank (lemmas and words), marked differences existed as well between (a) the proxy used for paradigmatics and the lexical measures, (b) the semantic fractions (major and minor) and the paradigmatic proxies and (c) the paradigmatic proxies and the lexical measures (inflected and uninflected lexemes). It is clear, then, that the language of the NT encodes unique data paradigmatically, syntagmatically and stratally. These findings underscore why a linguistically developed and annotated corpus is critical to computationally-based language study.

10.3.3 Model Implication Three: Operating with Our Feet to the Fire via Abduction and IBE

Experimental design theory assists us in a principled way to reduce both Type I and Type II error, not merely in our experiments, but in our hermeneutic—the causal assignments that

⁴¹ Rudman, “The State of Authorship Attribution Studies,” 361.

⁴² Argamon, “Scalability Issues in Authorship Attribution,” 95.

⁴³ Mealand, “Computers in New Testament Research,” 108.

⁴⁴ See Section 2.1.

⁴⁵ The two clause complex annotations were not compared.

always emerge out of hermeneutical reflections.⁴⁶ If we are to be as empirical as Sampson and Biber argue that we must be,⁴⁷ we must embrace experimental design. Moreover, if we are to be *truly* open to the diverse stances of others, especially if they do not share our own priors and perspectives,⁴⁸ then we submit that we must also embrace abduction as an important strand in terms of method. Abduction, as we have used it here, is a creative, poiemic, non-positivistic stance and praxis that allows us to embrace the propositions of others. This embrace is open, fulsome, and genuine—but it is not *fundamentalist*. It denies hegemonies and privilege from whatever quarter it may come, and it does so by being critical. Specifically, it transforms its own perspectives as well as the perspectives of others (to the extent possible) into testable hypotheses and then assays the adequacy of those hypotheses using IBE. The proposition that demonstrates itself to be repeatedly abductively inadequate cannot by abduction become *essentially deprived*. Rather it does become *inferentially unuseful* as an explanation for the phenomena in question. This is a cycle that I propose NT stylistics must embrace if it is to protect itself from its own hegemonies as well as the hegemonies of others.

10.3.4 Model Implication Four: Expanding the Automation Boundary

In Chapter Three I detailed study after study in which conclusions were propounded using inadequate linguistic measures with very thin sets of analysis. This led to two massive Kuhnian-caliber “messes” in computational stylistics.⁴⁹ I traced one reason to the menacing overlay of positivism.⁵⁰ But I submit there is another reason just as compelling: *inadequate test depth*. I argue that we can yield substantial breakthroughs in stylistics if we expand the

⁴⁶ See Section 2.4.

⁴⁷ Sampson, *Empirical Linguistics*, 22.

⁴⁸ Or perhaps, depending on how solipsistic one may be, our Kantian forms and structures.

⁴⁹ Rudman, “The State of Authorship Attribution Studies,” 351. Cf. also Juola, “Authorship Attribution for Electronic Documents,” 121. Regarding the second “mess” see Argamon, “Scalability Issues in Authorship Attribution,” 95.

⁵⁰ Section 3.8.

automation boundary in NT stylistics. Perhaps this sounds futuristic. It is not. Modern methods of automation and knowledge discovery have been developed and now automate knowledge discovery in commercial settings worldwide. These automation tools are software that run experiments demanded of them by an experimental design repository, and scan for important results. We built a set of automation tools for this study. Despite its rather hodge-podge nature, even this limited use of automation increased the number of experiments we were able to run about tenfold. This level of automation, even as primitive as it was, yielded a depth of computational stylistic analysis that to my knowledge has not been executed on the NT before.

Specifically, this study (again to the best of my knowledge):

- Explored more linguistic species in the NT than ever before: 51,199 linguistic categories drawn for the GNT (including paradigmatics, syntagmatics, and semantics)
- Performed more bivariate analyses in the NT than ever before: 11,110.⁵¹ Each of these was assessed with G&K's tau, and a suite of IT measures
- Performed more nonparametric analyses than ever before: 7,128 HLLA
- Performed (I believe) more CA and MCA analyses in one study than ever before: 270

Moreover, I developed database automation tools that allowed the rapid creation of a data repository that:

- Included all the major traditional grammatical annotations of the GNT
- Integrated stratal data (the sematic domains of Louw and Nida)
- Integrated discourse data (Lexham's GNT⁵²)
- Integrated SFL-based syntagmatic data (spanning from lemma to clause complex)
- Now contains over 2,000 separate linguistic probes (variables).

I am not, of course, arguing that quantity makes quality. But per the critique of Zipf, all linguistic species should be assayed,⁵³ and per the examples of Stratil and Oakely and Lana

⁵¹ In addition to the 1,210 tables mentioned in Section 6.2, this study explored these same nine measures by eleven theories (including a control group) by centiles.

⁵² Not used in this study.

⁵³ Hence he looked at all distributions of words from the least to the most frequent. Zipf, *Selected Studies of the Principle of Relative Frequency*, 22–27.

among others, multiple multivariate analyses are needed to gain multiple perspectives into the data.⁵⁴ Automation must be leveraged to those ends.

10.4 Directions for Further Research

A number of next steps are indicated to complete our experimental design regarding genre and authorship as well as to develop more definitive authorship findings. These steps include:

1. Broadening the language measures used—especially syntagmatically (i.e. with collocations, lexical priming, and discourse markers) and metafunctionally (such things as verbal aspect, participant chains, textual/logical relations such as cohesion and the like)—so as to continue to quantify their relation to genre and authorship.⁵⁵ In our view over 100 other linguistic measures which span both the metafunctions of language and the systems of choice natural to the Koine must be explored to more comprehensively characterize the dynamics of the relation between genre and authorship in the GNT.
2. Propose and test more genre constructs. These potentially may come from more linguistically-informed theoretical constructs (per Pitts) or from *a posteriori* analytics or some abductive combination of both.
3. Inspect for other covariates which may confound the interpretation of authorship, such as occasion alone and theological community (Hellenistic, Palestinian/Semitic) alone, etc.
4. Further partition the data to understand the covariance structure of the hotly debated low-frequency partition.

⁵⁴ Stratil and Oakley, “A Disputed Authorship Study of Two Plays,” 153–59; Lana, “Xenophon’s Athenaion Politeia,” *passim*.

⁵⁵ In particular, the clause complexes as measured were as a lineal set of function slots from the head term regardless of whether or not that complex included combinations of primary, secondary, and embedded clauses. Further work will respect *taxis* by analyzing clauses by type (primary or secondary).

5. Subtract from the total variance structure of the GNT the interactions between these covariates so that the different authorship theories can be tested apart from these confounding effects. This involves constructing derivative contingency tables by subtracting the contributions of genre from each cell and reanalyzing resulting contingency tables in terms of authorship.
6. Broaden the effect-level models from HLLA alone to Bayesian models, structural information theory models, and the like.⁵⁶ This will not only provide confirmation of the HLLA findings but address some of the limitations related to the model form of HLLA.⁵⁷
7. Set up tests of inference between authorship theories once genre and other confounding covariates have been subtracted.
8. Explore the resulting variance structure at the row level (e.g. the individual lemmas, words, etc.). In this way we can develop markers empirically rather than *ad hoc*.⁵⁸
9. Use Stirling numbers of the second kind,⁵⁹ and specifically the Bell number (a construct in combinatorics),⁶⁰ to generate and test all combinations of genre and authorship. These combinations will allow us to empirically rather than notionally uncover the “best” genre and authorship theories respectively. The most parsimonious set of these highest performing solutions will then be analyzed for commonalities which should provide us with persistent categories of genre and authorship that are necessary for a solution to

⁵⁶ Streitberg, “Exploring Interactions in High-Dimensional Tables,” *passim*.

⁵⁷ See the discussion by Krippendorff, *Information Theory*, 89.

⁵⁸ Other second tier work would include: (1) Breaking out clauses by primary and secondary clauses. Currently clauses are just seen as a linear sequence of slots without distinguishing levels of taxis. (2) Using exact tests for further exploring low-frequency forms of the linguistic measures. (3) Expressing each form in words per 1,000. This “de-weights” the frequency of the words by counting every word with an equal weight regardless of its frequency distribution. (4) Performing measures of association of standardized tables per Reynolds, *Analysis of Nominal Data*, 70.

⁵⁹ Stirling numbers of the second kind sum the total number of ways to partition a set of n -labeled objects (i.e. starting groupings of texts) into k nonempty unlabeled groups (Knuth, *The Art of Computer Programming*, 65).

⁶⁰ Bell numbers are the sum of Stirling numbers of the second kind.

“work.” By so doing, synchronic analysis of the GNT may be able to provide NT scholars with new insight into long-standing matters of occasion and introduction in the GNT.

Enhancing the IGNT, the first step required to execute these nine further research activities, is currently underway.

10.5 A Final Hermeneutical Reflection

In one sense I hope to have made clear that what I have attempted here is certainly not “just running math” on the New Testament. That is precisely what has been tried and has failed. The question I have tried to raise in this work is quintessentially a hermeneutical one. I hope that I have also been clear that I am not arguing that uncritical pre-modern conservative/pietistic hermeneutics on the one hand or materialist enlightenment hermeneutics on the other should be dragged back to the hermeneutical table. Neither dogma nor positivism has gotten us very far. What can computational stylistics offer the hermeneut? Nothing, surely, if it is done positivistically. If Bockmuehl’s assessment is correct, ours is a discipline in “disarray,” “fractured and aimless.”⁶¹ According to Watson, moreover, methodological pluralism and ever-narrowing specialization have resulted in an ongoing process of fragmentation.⁶² Aichele, Miscall, and Walsh agree, viewing the NT landscape as a one of “entrenched camps” and “embattled groups...who speak to one another...only in sniping, intellectually unengaged footnotes.”⁶³ What *does* one do with an academic discipline, asks Robert W. Jenson, which has almost done away with its own object of study?⁶⁴ The landscape we find ourselves in,⁶⁵ if we can

⁶¹ Bockmuehl, *Seeing the Word*, 30–39.

⁶² Watson, *Text and Truth*, 26.

⁶³ Aichele et al., “Historical-Critical and Postmodern Interpretations of the Bible,” 383.

⁶⁴ Endorsement to Bockmuehl, *Seeing the Word*. Adam paints the same picture of biblical theology, asking “Can these bones live?” Adam, *Faithful Interpretation*, 19.

⁶⁵ A phrase adapted from the prolific, emergent lay-theologian, Brian McLaren; McLaren, *The Story We Find Ourselves In*.

fine-tune the metaphor of Aichele et al., is one of widely separated clans of NT scholars warming themselves around their own campfires fuelled by their own criteria and methods. They are aware of other campfires, to be sure, but for the most part those fires burn only upon the distant horizon, yielding neither heat nor light amenable to their own concerns. The crux of the matter is that NT scholars even disagree as to whether fragmentation is evidence of an underlying problem or evidence of the hard-fought solution to the problem. One could, for instance, cull opinions only from the former group, recognized lights in the NT firmament such as Green,⁶⁶ Thiselton,⁶⁷ Fowl,⁶⁸ and Jeanrond.⁶⁹ While this cadre of scholars are themselves by no means cut from the same hermeneutical cloth, they are united at least in the conviction that unbounded and untethered plurality leads only to hermeneutical vacuity. For a substantial and apparently growing number of scholars within our guild, however, fragmentation and pastiche, methodological dissonance and scholarly hermitage are not the problem but the long-awaited deliverance from the problem.⁷⁰ These, with some qualification, would include scholars such as Moore,⁷¹ Tracy,⁷² Adam,⁷³ Aichele⁷⁴ and Sherwood.⁷⁵ For these scholars, no fundamental rupture of the biblical-scholarly *épistémè* has actually yet occurred, or at least not one comparable to the one that brought our discipline into being. In their view the modernist New Testament ship of

⁶⁶ See especially Green, "Scripture and Theology," 23–43. See also Green, ed., *Hearing the New Testament*; Green, *Practicing Theological Interpretation*.

⁶⁷ Thiselton, *New Horizons in Hermeneutics*; Thiselton, *Interpreting God and the Postmodern Self*; Thiselton, *Thiselton on Hermeneutics*; Thiselton and Marko, "Thirty Years of Hermeneutics," passim; Porter and Malcolm, eds., *Horizons in Hermeneutics*.

⁶⁸ Fowl, *The Theological Interpretation of Scripture*, passim; Fowl, *Engaging Scripture*.

⁶⁹ Jeanrond, *Text and Interpretation*; Jeanrond, *Theological Hermeneutics*; Jeanrond, "Theology in the Context of Pluralism and Postmodernity: David Tracy's Theological Method," passim.

⁷⁰ Fowl has a different perspective on this same phenomena, asserting "determinate," "antideterminate" and "undeterminate" interpretations of Scripture. Fowl, *Engaging Scripture*, 32–61. "No Solution" scholars tend toward the latter two categories.

⁷¹ For a compendium of his essays see Moore, *The Bible in Theory*.

⁷² Tracy, *Plurality and Ambiguity*.

⁷³ Adam, *Handbook of Postmodern Biblical Interpretation*; Adam, *Faithful Interpretation*.

⁷⁴ Aichele et al., "Historical-Critical and Postmodern Interpretations of the Bible," passim.

⁷⁵ Moore and Sherwood, "Biblical Studies 'After' Theory: (Part 2)," 90.

state—to shift the metaphor—has been merely dented by some well-placed postmodern salvos. What needs to be done, fully and finally, is to finish the job. Only when the metanarrative-freighted superstructure of the current New Testament scholarly *épistémè* is rent asunder and slips beneath the waves will it take with it the rotting ballast of Enlightenment/imperial/colonial/critical privilege that it has carried far too long.

What I have argued for here, then, consistent with the way opened by Kripke (contingent necessary truths) and Peirce, is that one exploratory way forward from the current impasse is to expand the empirical phenomenal boundary toward us (the subject) again, *but this time nonpositivistically*. In this way, I believe, a chastened and humbled “empirical turn” can be added to the linguistic and sociological turns already underway in NT studies. The framework of such a turn is abductive, as I have argued in Chapter Two. What I envision is a kind of hermeneutical spiral, but different from Osborne’s. The first element is abduction itself, the use of inference and validation to arrive at the best explanation. This is what I have tried here—empiricism to be sure, but a flavor of it extracted from its damning association with positivism. The second is historical analysis with openness to possibilities. A hermeneutic of openness requires that new stances in terms of history and historicizing be explored. This is not to say that work in the historical-critical venue should be in any way devalued. Yet it, along with all such methods used in NT studies must ask of itself what it should ask of all methods: Are our criteria testable? Are they open to falsification? If the answers to both questions are in the negative, then from what quarter can such a claim to privilege come? The third element is *textual* in two ways: (a) via synchronic analysis of the text but one that is linguistically-informed (Tracy’s linguistic

turn)⁷⁶ and (b) via responsible diachronic textual analysis that speaks into what actually can be said rather than speculated regarding the text's prior state. The fourth is openness to existential data—the inbreaking of the text and the Person behind the text into our *Erklärung*. These four poles, the empirico-critical pole, the historical pole (as distinguished from but not exclusive of the historical-critical pole), the textual pole and the existential pole are all needful for us to develop a hermeneutic which remains open to (does not deprivilege) the claims of the text before we even begin. Here then is Luz's core idea – but not his assignment of causes.⁷⁷ In agreement with Luz, NT hermeneutics, and to a large degree biblical hermeneutics, has failed not because, *ipso facto*, it should not have a place at the larger table of social and philosophical discourse. It has failed because it has been judged irrelevant to it. In two of its major streams today, for instance, NT hermeneutics is ahistorical and mystical in a world whose needs are very physical and where spiritual realities are not limited to ecstasies. NT hermeneutics, per Luz, has indeed failed, and it has failed because understandings of the Bible and not the Bible itself have formed the foundation of our theologies. The four poles of hermeneutics just presented, each one stripped of its speculations and its own totalizing tendencies, will, I submit, go far toward developing a post-postmodern hermeneutic.

⁷⁶ Tracy, *Plurality and Ambiguity*, 47. This is Tracy's phrase, although it has been a slow turn indeed. Most NT scholars, I think it safe to say, would not number linguistics among their primary or even secondary research interests.

⁷⁷ I refer here the earlier Ulrich Luz and his brilliant defence of reception history in his 1998 SNTS presidential address. To the best of my knowledge this is Luz's most extensive interaction with the "problem" in NT studies. Luz, *Studies in Matthew*, 313–32.

Appendix

The Measurement, Verificational and Hermeneutical Dimensions of the Experimental Design

(From the Design Implications in Chapter 3)

Code	Criteria	3	4	5	6	7	8	9	10	11	12	13
Measurement Dimension (What to Measure)												
<i>Textual Scope</i>												
MTK	Test all the extant Koine		b						l			
MTG	Test within the Greek New Testament		c									
MTB	Test text subsets between Koine and GNT											
<i>Linguistic Model (Annotation to Use)</i>												
MLH	Use Halliday's SFL as the primary model		a									b
MLL	Use Lamb's SG as a model											c
MLD	Use Dik's Functional Grammar											c
MPL	Use Pike's Tagmemics as a model											c
MLA	Use SIL (South African) as a model											c
<i>Linguistic Probes (Where to Annotate)</i>												
MPS	For each strata (e.g. semantics)	d	d	f	b	a,e	e,g,k	d	i,n	a,c		a
MPP	For each paradigm of language				b	a,e	e,g,k	d	i,n	a,c		a
MPR	For each rank of language				b	a,e	e,g,k	d	i,n	a,c,f		a
MPM	For each (meta) function of language				b	a,e	e,g,k	d	i,n	a,c		a
MPD	For various discourse structures				d	a,e	e,g,k	d	i,n	a,c		a
MPI	For information flow				f							
MPF	For each syntax level (Fawcett)											
MPC	Explore measures separately before combining then into higher order categories					b,h						
<i>Features to Select</i>												
MFC	Use all tokens w/in each language layer			a		g	p		a,e,h	b,e		
MFV	Use most variant tokens within a layer					i			h,j,k	B		
MFR	Use most representative tokens / layer					m			h,j,k	B		
MFH	Explore low-/high-frequency bins					k		a	h,j,k	B		
MFT	Explore top and bottom N by frequency	a,c,p				k		a	h,j,k	B		
Verificational Dimension (How to Validate)												
VMF	Use Fisherian (Classical) Statistics			b		d,l						
VMB	Use Bayesian Inference			c		d,l						
VMO	Use Bootstrapping-Based Inference			d		d,l						
VMM	Use Mixed Methods of Inference			e		d,l						
VMC	Use combinations of the above if possible					d,l						
VMH	Use multiple hold-back strategies for SMVA											
VMT	Triangulate findings with diachronic approaches											
Hermeneutical Dimension (What Frameworks to Assign to Stylistic Significance)												
HI	Idiolectic (Author, Amanuensis, Pastiche)						b,d,f					
HSD	Sociolectic (Genre, Context, Audience, etc.)						l					
HD	Dialectic or Diglossic											
HST	Stylochometric											
HO	Other											

Table A-1.

The Initial Dimensions of an Experimental Design (Grouping and Analytical Dimensions)

Code	Criteria (subsections in chapter three)	3	4	5	6	7	8	9	10	11	12	13
Grouping Dimension												
What A Priori Hypotheses to Test (Groups)												
GHA	Test authorship hypotheses vs. rest.					f				g		
GHR	Test genre hypotheses vs. rest	b				f	b,d			g		
GHG	Test register hypotheses vs. rest					f	b,d,o			g,m		
GHD	Test date hypotheses vs. rest	e,m				f	b,d			g		
GHS	Test subject matter hypotheses vs. rest					f	b,d			g		
GHP	Test audience plurality vs. rest					f	b,d,o			g		
GHC	Test context of situation vs. rest		e			f				g		
Analytical Dimension												
How to Select Samples												
ASE	Equal size samples where required	f						n			d	
ASD	Sufficient degrees of freedom	g,i						i,n			d	
ASF	Sufficient number of features	h									d	
How to Represent the Data												
ARR	Valid selection of representative data											
ARV	Valid visual display of quantitative data						j					
Features to Select												
MFC	Use all tokens within each language layer			a		g	p		a,e,h	b,e		
MFV	Use most variant tokens within a layer					i			h,j,k	b		
MFR	Use most representative tokens / layer					m			h,j,k	b		
MFF	Explore low/high-frequency bins					k		a	h,j,k	b		
MFT	Explore top and bottom N by frequency	a,c,p				k		a	h,j,k	b		
What Text Grouping to Test												
ATP	Test with parametric methods (e.g. t-tests)											
ATN	Test with non-paramet. methods (e.g. χ^2)											
ATI	Test with Info Theory-based methods											
ATB	Test with Bayesian											
What Univariate Tests to Run												
AUP	Parametric methods (e.g. t-tests)											
AUI	Test with Info. Theory-based methods						m					
AUB	Test with Bayesian											
What Bivariate Tests to Run												
ABA	Association (χ^2 , Cramér's V)											
ABP	Direction (PRE methods)											
ABI	Information (Entropy, UC)											
Multivariate Description and Inference												
AMP	Parametric (e.g., MANCOVA)					c						
AMN	Nonparametric (e.g. HLLA, GLL)					c						
AMI	Information-Theory-Based					c						
AMB	Bayesian					c						
Multivariate Extractive Tests to Run												
AEP	Parametric (e.g. PCA,FA)		f			c			b,e			
AEN	Nonparametric (e.g. CA,MCA,CATPCA)		f			c			e			
AEM	Mixed Parametric (e.g. ALSOS)		f			c			e			
Multivariate Causal (Latent) Tests to Run												
ACP	Parametric (e.g. SEM,LISREL)	j			e	c			g			
ACN	Nonparametric (e.g. LDA)				e	c			g			
ACI	Information-Theory-Based				e	c			g			
ACG	Graphical models				e	c			g			
Multivariate Supervised (SMVA) Test to Run												
ASP	Parametric (e.g. MDA)					j			f			
ASM	Machine Learning (NN, SVM, etc.)					j			f			
ASI	Information-Retrieval-Based					j			f			
ASE	Ensemble-Based					j			f			
Avoid Common Sources of Error												
AES	Avoid scalar distribution parameters	k,l										
AEC	Link to a cause using causal methods	n,o										
AED	Use the entire distribution or bins			a,c			a,c		b,d			
AEE	Control for confounding effects							c				
AES	Run sup. MVA on EMVA components								p	e,h		
AER	Meet all MV pre-requisites								f			
AEM	Use multiple independent MVA techniques								c,o			
AEF	Ensure MVA's have high fit								r			
AEO	Implement research with adequate controls											
Improve Productivity												
APS	Via Software automation									g		
APH	Via Hardware optimization											

Table 2.

Bibliography

- Aageson, James W. "The Pastoral Epistles, Apostolic Authority, and the Development of the Pauline Scriptures." In *The Pauline Canon*, edited by Stanley E. Porter, 5–26. Pauline Studies (PAST) 1. Leiden; Boston: Brill, 2004.
- . *Paul, the Pastoral Epistles, and the Early Church*. Library of Pauline Studies 2. Peabody, MA: Hendrickson, 2008.
- Abbasi, Ahmed, and Hsinchun Chen. "Applying Authorship Analysis to Extremist-Group Web Forum Messages." *Intelligent Systems, IEEE* 20 (2005) 67–75.
- . "Visualizing Authorship for Identification." In *Intelligence and Security Informatics*, edited by Sharad Mehrotra et al., 60–71. Lecture Notes in Computer Science 3975. Berlin; Heidelberg: Springer, 2006.
- . "Writeprints: A Stylometric Approach to Identity-Level Identification and Similarity Detection in Cyberspace." *ACM Transactions on Information Systems (TOIS)* 26 (2008) 1–29.
- Abbasi, Ahmed, et al. "Sentiment Analysis in Multiple Languages." *ACM Transactions on Information Systems (TOIS)* 26 (2008) 1–34.
- . "Stylometric Identification in Electronic Markets: Scalability and Robustness." *Journal of Management Information Systems* 25 (2008) 49–78.
- Achtemeier, Paul J. "Omne verbum sonat: The New Testament and the Oral Environment of Late Western Antiquity." *Journal of Biblical Literature* 109 (1990) 3–27.
- Achtemeier, Paul J., et al. *Introducing the New Testament: Its Literature and Theology*. Grand Rapids: Eerdmans, 2001.
- Adam, A.K.M. *Faithful Interpretation: Reading the Bible in a Postmodern World*. Minneapolis: Fortress, 2006.
- . *Handbook of Postmodern Biblical Interpretation*. St. Louis: Chalice, 2000.
- Agresti, Alan. *An Introduction to Categorical Data Analysis*. Wiley Series in Probability and Statistics 11. New York: Wiley, 1996.
- . "A Survey of Exact Inference for Contingency Tables." *Statistical Science* 7 (1992) 131–53.
- Aichele, George, et al. "An Elephant in the Room: Historical-Critical and Postmodern Interpretations of the Bible." *Journal of Biblical Literature* 128 (2009) 383–404.

- Akaike, Hirotugu. "Prediction and Entropy." In *Selected Papers of Hirotugu Akaike*, 387–410. New York; Berlin; Heidelberg: Springer, 1998.
- Aldenderfer, Mark S., and Roger K. Blashfield. *Cluster Analysis*. Quantitative Applications in the Social Sciences 44. Newbury Park: Sage, 1984.
- Allen, John R. "Methods of Author Identification through Stylistic Analysis." *The French Review* 47 (1974) 904–16.
- Alviar, J. J. "Recent Advances in Computational Linguistics and Their Application to Biblical Studies." *New Testament Studies* 54 (2008) 139–59.
- Amador, J.D.H. *Academic Constraints in Rhetorical Criticism of the New Testament: An Introduction to a Rhetoric of Power*. Library of New Testament Studies 174. Sheffield: Sheffield Academic, 1999.
- Argamon, Shlomo. "Scalability Issues in Authorship Attribution." *Literary and Linguistic Computing* 27 (2012) 95–97.
- Argamon, Shlomo, and Jeff Dodick. "Corpus-Based Study of Scientific Methodology: Comparing the Historical and Experimental Sciences." In *Computing Attitude and Affect in Text: Theory and Applications*, edited by James G. Shanahan et al., 215–31. The Information Retrieval Series 20. Dordrecht: Springer, 2006.
- Argamon, Shlomo, and Moshe Koppel. "The Rest of the Story: Finding Meaning in Stylistic Variation." In *The Structure of Style*, edited by Shlomo Argamon et al., 79–112. Berlin; Heidelberg: Springer 2010.
- Argamon, Shlomo, et al. "Automatically Profiling the Author of an Anonymous Text." *Communications of the ACM* 52 (2009) 119–23.
- . "Gender, Genre, and Writing Style in Formal Written Texts." *Interdisciplinary Journal for the Study of Discourse* 23 (2003) 321–46.
- . "Language Use Reflects Scientific Methodology: A Corpus-Based Study of Peer-Reviewed Journal Articles." *Scientometrics* 75 (2008) 203–38.
- . "Style Mining of Electronic Messages for Multiple Authorship Discrimination: First Results." Paper presented at the Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, August 24th–27th, 2003.
- . "Stylistic Text Classification Using Functional Lexical Features." *Journal of the American Society for Information Science and Technology* 58 (2007) 802–22.

- Arppe, Antti. "Univariate, Bivariate, and Multivariate Methods in Corpus-Based Lexicography: A Study of Synonymy." Ph.D. diss., University of Helsinki, 2008.
- Atkinson, A. C. *Optimum Experimental Designs, with SAS*. Oxford Statistical Science Series 34. Oxford: Oxford University Press, 2007.
- Aune, David Edward. *The New Testament in its Literary Environment*. Library of Early Christianity 8. Philadelphia: Westminster, 1987.
- . "What is Gospel?" In *Gospel Perspectives: Studies of History and Tradition in the Four Gospels*, edited by R. T. France and David Wenham, 9–60. Sheffield: JSOT, 1981.
- Baayen, R. Harald. *Word Frequency Distributions*. Dordrecht; Boston: Kluwer Academic, 2001.
- Baayen, R. Harald, et al. "An Experiment in Authorship Attribution." Paper presented at the 6th JADT, Saint-Malo, March 13th-15th, 2002.
- . "Outside the Cave of Shadows: Using Syntactic Annotation to Enhance Authorship Attribution." *Literary and Linguistic Computing* 11 (1996) 121–31.
- Baird, J. Arthur. "Content Analysis and the Computer: A Case-Study in the Application of the Scientific Method to Biblical Research." *Journal of Biblical Literature* 95 (1976) 255–76.
- Baker, Paul, et al. *A Glossary of Corpus Linguistics*. Edinburgh: Edinburgh University Press, 2006.
- Bakhtin, M.M. *Speech Genres and Other Late Essays*. Translated by Vern W. McGee. Slavic Series 8. Austin: University of Texas Press, 1986.
- Barr, George K. "A Computer Model for the Pauline Epistles." *Literary and Linguistic Computing* 16 (2001) 233–50.
- . "Contrasts in Scale and Genre in the Letters of Paul and Seneca." *Irish Biblical Studies* 18 (1996) 16–25.
- . "Graphical Analysis of the Sentence Length Distribution Curve and Non-rational Components." *Literary and Linguistic Computing* 16 (2001) 375–88.
- . "The Impact of Scalometry on New Testament Letters." *The Expository Times* 114 (2002) 3–9.
- . "Interpolations, Pseudographs, and the New Testament Epistles." *Literary and Linguistic Computing* 17 (2002) 439–55.

- . “Literary Dependence in the New Testament Epistles.” *Irish Biblical Studies* 19 (1997) 148–60.
- . “Scalometry and the Dating of the New Testament Epistles.” *Irish Biblical Studies* 22 (2000) 71–90.
- . *Scalometry and the Pauline Epistles*. The Library of New Testament Studies 261. London; New York: T & T Clark, 2004.
- . “Two Styles in the New Testament Epistles.” *Literary & Linguistic Computing* 18 (2003) 235–48.
- . “The Use of Cumulative Sum Graphs in Literary Scalometry.” *Literary & Linguistic Computing* 12 (1997) 103–11.
- Barr, James. *The Semantics of Biblical Language*. London: Oxford University Press, 1961.
- Barrett, C. K. “A Critical and Exegetical Commentary on the Pastoral Epistles.” *Journal of Theological Studies* 52 (2001) 824–27.
- Barrón-Cedeño, Alberto, et al. “Plagiarism Detection across Distant Language Pairs.” In *Proceedings of the 23rd International Conference on Computational Linguistics*, 37–45. Beijing: Association for Computational Linguistics, 2010.
- Barthes, Roland. *Image, Music, Text*. Translated by Stephen Heath. Fontana Communication Series 14. Glasgow: Fontana Collins, 1977.
- Bartholomew, D. J. “Probability, Statistics and Theology.” *Journal of the Royal Statistical Society* 151 (1988) 137–78.
- Barton, John. *The Biblical World*. London; New York: Routledge, 2004.
- Bauckham, Richard. “Pseudo-Apostolic Letters.” *Journal of Biblical Literature* 107 (1988) 469–94.
- Baum, Armin Daniel. “Authorship and Pseudepigraphy in Early Christian literature: A Translation of the Most Important Source Texts and an Annotated Bibliography.” In *Authorship and Pseudepigraphy in Early Christian Literature*, edited by Stanley E. Porter and Gregory P. Fewster, 11–64. Pauline studies 8. Leiden; Boston: Brill, 2013.
- . “Literarische Echtheit als Kanonkriterium in der alten Kirche.” *Zeitschrift für die neutestamentliche Wissenschaft* 88 (1997) 97–110.
- Baur, Ferdinand Christian. *Die sogenannten Pastoralbriefe des Apostels Paulus*. Stuttgart: J.G. Cotta, 1835.

- . *Paulus, der Apostel Jesu Christi: Sein Leben und Wirken, seine Briefe und seine Lehre*. Stuttgart: Becher & Müller, 1845.
- Beatrice, Pier F. “Forgery, Propaganda and Power in Christian Antiquity.” In *Alvarium: Festschrift für Christian Gnilka*, edited by Wilhelm Blümer et al., 39–51. *Jahrbuch für Antike und Christentum* 33. Münster, Westfalen: Aschendorff, 2002.
- Becker, Mark P. “Models for the Analysis of Association in Multivariate Contingency Tables.” *Journal of the American Statistical Association* 84 (1989) 1014–19.
- Becking, Bob. “Bee’s Dating Formula in the Book of Nahum.” *Journal for the Study of the Old Testament* 18 (1980) 100–04.
- Bee, Ronald E. “Statistical Methods in the Study of the Masoretic Text of the Old Testament.” *Journal of the Royal Statistical Society* 134 (1971) 611–22.
- . “A Statistical Study of the Sinai Pericope.” *Journal of the Royal Statistical Society* 135 (1972) 406–21.
- . “The Use of Statistical Methods in Old Testament Studies.” *Vetus Testamentum* 23 (1973) 257–72.
- Bell, Allan. “Language Style as Audience Design.” *Language in Society* 13 (1984) 145–204.
- Bennett, Paul E. “The Statistical Measurement of a Stylistic Trait in Julius Caesar and As You Like It.” *Shakespeare Quarterly* 8 (1957) 33–50.
- Benzecri, J. P. “Elaboration statistique de données sur ordinateur; Application à l’analyse des textes; contributions attendues à l’étude de la Bible.” *Cahiers de l’Analyse des Données* 11 (1986) 7–30.
- Berger, Helmut, and Dieter Merkl. “A Comparison of Text-Categorization Methods Applied to N-Gram Frequency Statistics.” In *AI 2004: Advances in Artificial Intelligence*, edited by Geoffrey Webb and Xinghuo Yu, 998–1003. Lecture Notes in Computer Science 3339. Berlin; Heidelberg: Springer, 2005.
- Berlyne, Daniel E. *Aesthetics and Psychobiology*. New York: Appleton-Century-Crofts, 1971.
- Berry, M. W., et al. “Document Classification Using Nonnegative Matrix Factorization and Underapproximation.” In *ISCAS: 2009 IEEE International Symposium on Circuits and Systems*, 2782–85. New York: IEEE Xplore, 2009.
- Betz, Hans Dieter. *Galatians: A Commentary on Paul’s Letter to the Churches in Galatia*. Philadelphia: Fortress, 1979.

- Biber, Douglas. *Dimensions of Register Variation: A Cross-Linguistic Comparison*. Cambridge; New York: Cambridge University Press, 1995.
- . “Investigating Macroscopic Textual Variation Through Multifeature/Multidimensional Analyses.” *Linguistics* 23 (1985) 337–60.
- . “The Multi-Dimensional Approach to Linguistic Analyses of Genre Variation: An Overview of Methodology and Findings.” *Computers and the Humanities* 26 (1992) 331–45.
- . *University Language: A Corpus-Based Study of Spoken and Written Registers*. Studies in Corpus Linguistics 23. Amsterdam; Philadelphia: John Benjamins, 2006.
- . *Variation Across Speech and Writing*. Cambridge; New York: Cambridge University Press, 1988.
- Biber, Douglas, and Edward Finegan. *Sociolinguistic Perspectives on Register*. New York; Oxford: Oxford University Press, 1994.
- Biber, Douglas, et al. *Discourse on the Move: Using Corpus Analysis to Describe Discourse Structure*. Studies in Corpus Linguistics 28. Amsterdam; Philadelphia: John Benjamins, 2007.
- . *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge; New York: Cambridge University Press, 1998.
- Binongo, José Nilo G. “Who Wrote the 15th Book of Oz? An Application of Multivariate Analysis to Authorship Attribution.” *Chance* 16 (2003) 9–17.
- Binongo, José Nilo G., and M. W. A. Smith. “The Application of Principal Component Analysis to Stylometry.” *Literary and Linguistic Computing* 14 (1999) 445–65.
- . “A Bridge between Statistics and Literature: The Graphs of Oscar Wilde’s Literary Genres.” *Journal of Applied Statistics* 26 (1999) 781–87.
- . “Project Jacobean: A Mathematical Exploration of a Literary Era.” *The Mathematics Teacher* 93 (2000) 338–44.
- . “Statistical Approaches to Philippine Literature.” *Philippine Studies* 45 (1997) 500–38.
- Bishop, Yvonne. M., et al. *Discrete Multivariate Analysis: Theory and Practice*. New York: Springer, 2007.
- Black, C. Clifton. “Rhetorical Criticism and the New Testament.” *Proceedings of the Eastern Great Lakes and Midwest Biblical Societies* 8 (1988) 77–92.

- Black, David Alan. *Linguistics for Students of New Testament Greek: A Survey of Basic Concepts and Applications*. 2nd ed. Grand Rapids: Baker, 1995.
- Black, David Alan, et al. *Linguistics and New Testament Interpretation: Essays on Discourse Analysis*. Nashville, TN: Broadman, 1992.
- Black, Stephanie L. *Sentence Conjunction in the Gospel of Matthew: Kai, De, Tote, Gar, Oun and Asyndeton in Narrative Discourse*. *Journal for the Study of the New Testament*. 216. Sheffield: Sheffield Academic, 2002.
- Blackith, Robert. "A Multivariate Analysis of Latin Elegiac Verse." *Language and Speech* 6 (1963) 196–205.
- Blei, David M., et al. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3 (2003) 993–1022.
- Blomberg, Craig L. "The Diversity of Literary Genres in the New Testament." In *New Testament Criticism & Interpretation*, edited by David Alan Black and David S. Dockery, 505–32. Grand Rapids: Zondervan, 1991.
- . *The Historical Reliability of the Gospels*. 2nd ed. Nottingham: Apollos, 2007.
- . "New Testament Genre Criticism for the 1990's." *Themelios* 15 (1990) 40–49.
- Blomberg, Craig L., and Jennifer Foutz Markley. *A Handbook of New Testament Exegesis*. Grand Rapids: Baker, 2010.
- Blum, Avrim L., and Pat Langley. "Selection of Relevant Features and Examples in Machine Learning." *Artificial Intelligence* 97 (1997) 245–71.
- Bockmuehl, Markus. *Seeing the Word: Refocusing New Testament Study*. Grand Rapids: Baker Academic, 2006.
- Boer, Roland. *Bakhtin and Genre Theory in Biblical Studies*. *Semeia* 63. Atlanta: Society of Biblical Literature, 2007.
- Bond, L. Susan. "1 Timothy 1:3:17." *Interpretation* 60 (2006) 314–17.
- Bopp, Franz. *A Comparative Grammar of the Sanskrit, Zend, Greek, Latin, Lithuanian, Gothic, German, and Slavonic Languages*. Translated by Edward Backhouse Eastwick. 2nd ed. London; Edinburgh: Williams and Norgate, 1856.
- Boreland, Hellen, and Patricia Galloway. "Authorship, Discrimination, and Clustering: Timoneda, Montesino, and Two Anonymous Poems." *Bulletin of the Association for Literary and Linguistic Computing* 8 (1980) 125–51.

- Boring, M. Eugene *An Introduction to the New Testament: History, Literature, Theology*. Louisville: Westminster John Knox, 2012.
- Botha, Antoine, et al. *Coping with Continuous Change in the Business Environment: Knowledge Management and Knowledge Management Technology*. Oxford: Chandos, 2008.
- Bourque, Linda Brookover. *Processing Data: The Survey Example*. Quantitative Applications in the Social Sciences 85. Newbury Park: Sage, 1992.
- Boyer, James L. "Project Gramcord: A Report." *Grace Theological Journal* 1 (1980) 97–99.
- Brainerd, Barron. "On the Distinction between a Novel and a Romance: A Discriminant Analysis." *Computers and the Humanities* 7 (1973) 259–70.
- . "On the Relation between Types and Tokens in Literary Text." *Journal of Applied Probability* 9 (1972) 507–18.
- . "Pronouns and Genre in Shakespeare's Drama." *Computers and the Humanities* 13 (1979) 3–16.
- . "Some Elaborations upon Gani's Model for the Type-Token Relationship." *Journal of Applied Probability* 18 (1981) 452–60.
- Brainerd, Barron, and Sun Man Chang. "Number of Occurrences in Two-State Markov Chains, with an Application in Linguistics." *The Canadian Journal of Statistics / La Revue Canadienne de Statistique* 10 (1982) 225–31.
- Bray, James H. *Multivariate Analysis of Variance*. Edited by Scott E. Maxwell. Quantitative Applications in the Social Sciences 54. Newbury Park: Sage, 1985.
- Brennan, Michael, et al. "Adversarial Stylometry: Circumventing Authorship Recognition to Preserve Privacy and Anonymity." *Acm Transactions on Information and System Security* 15 (2012).
- Brinegar, Claude S. "Mark Twain and the Quintus Curtius Snodgrass Letters: A Statistical Test of Authorship." *Journal of the American Statistical Association* 58 (1963) 85–96.
- Brockner, Joel, et al. "Culture and Procedural Justice: The Influence of Power Distance on Reactions to Voice." *Journal of Experimental Social Psychology* 37 (2001) 300–15.
- Brown, Milton P. *The Authentic Writings of Ignatius: A Study of Linguistic Criteria*. Durham, NC: Duke University Press, 1963.
- Brown, Steven R., and Lawrence E. Melamed. *Experimental Design and Analysis*. Quantitative Applications in the Social Sciences 74. Newbury Park: Sage, 1990.

- Brox, Norbert. *Falsche Verfasserangaben: Zur Erklärung der frühchristlichen Pseudepigraphie*. *Stuttgarter Bibelstudien* 79. Stuttgart: Katholisches Bibelwerk, 1975.
- Brueggemann, Walter. *Texts Under Negotiation: The Bible and Postmodern Imagination*. Minneapolis: Fortress, 1993.
- Brugmann, Karl. *A Comparative Grammar of the Indo-Germanic Languages: A Concise Exposition of the History of Sanskrit, Old Iranian (Avestic and Old Persian), Old Armenian, Greek, Latin, Umbro-Samnitic, Old Irish, Gothic, Old High German, Lithuanian and Old Church Slavonic*. Translated by R. Seymour Conway and W. H. D. Rouse. New York: Westermann & Company, 1891.
- Brunet, Étienne. "Le vocabulaire religieux dans trois siècles de littérature Française." In *Bible et Informatique : Interprétation, Herméneutique, Compétence Informatique*, 147–65. Universität Tübingen: Champin; Slatkine, 1991.
- Buch, Kai Rander. "A Note on Sentence-Length as Random Variable." In *Comptes rendus du onzième Congrès des mathématiciens scandinaves tenu à Trondheim 22-25 août 1949*, 272–75. Oslo: J. Grundt Tanums, 1952.
- Bunt, Harry C., and William J. Black. *Abduction, Belief, and Context in Dialogue: Studies in Computational Pragmatics*. Natural Language Processing 1. Amsterdam; Philadelphia: J. Benjamins, 2000.
- Burridge, Richard A. *What are the Gospels?: A Comparison with Graeco-Roman Biography*. Society for New Testament Studies 70. Cambridge; New York: Cambridge University Press, 1992.
- Burrows, John F. *Computation into Criticism: A Study of Jane Austen's Novels and an Experiment in Method*. Oxford: Clarendon, 1987.
- . "Not Unless You Ask Nicely: The Interpretative Nexus between Analysis and Information." *Literary & Linguistic Computing* 7 (1992) 91–109.
- . "'An Ocean Where Each Kind. . .': Statistical Analysis and Some Major Determinants of Literary Style." *Computers and the Humanities* 23 (1989) 309–21.
- . "Questions of Authorship: Attribution and Beyond: A Lecture Delivered on the Occasion of the Roberto Busa Award ACH-ALLC 2001, New York." *Computers and the Humanities* 37 (2003) 5–32.
- . "A Second Opinion on 'Shakespeare and Authorship Studies in the Twenty-First Century'." *Shakespeare Quarterly* 63 (2012) 355–92.
- . "Word-Patterns and Story-Shapes: The Statistical Analysis of Narrative Style." *Literary & Linguistic Computing* 2 (1987) 61–70.

- Burrows, John F., and D. Hugh Craig. "Authors and Characters." *English Studies* 93 (2012) 292–309.
- . "Lyrical Drama and the "Turbid Mountebanks": Styles of Dialogue in Romantic and Renaissance Tragedy." *Computers and the Humanities* 28 (1994) 63–86.
- Butler, Christopher. *Systemic Linguistics: Theory and Applications*. London: Batsford Academic and Educational, 1985.
- Cadbury, Henry J. "Some Foibles of New Testament Scholarship." *Journal of Bible and Religion* 26 (1958) 213–16.
- Calle-Martin, J., and A. Miranda-Garcia. "Stylometry and Authorship Attribution: Introduction to the Special Issue." *English Studies* 93 (2012) 251–58.
- Campbell, Gordon, et al. "The Provenance of *De Doctrina Christiana*." *Milton Quarterly* 31 (1997) 67–117.
- Campbell, Lewis. *The Sophistes and Politicus of Plato*. Oxford: Clarendon, 1867.
- Can, Fazli, and Jon M. Patton. "Change of Writing Style with Time." *Computers and the Humanities* 38 (2004) 61–82.
- Candlish, J.S. "On the Moral Character of Pseudonymous Books: II." *The Expository Times* 4 262–79.
- Carnap, Rudolf. "On the Character of Philosophic Problems." *Philosophy of Science* 51 (1984) 5–19.
- Carroll, John B. "Vectors of Prose Style." In *Style in Language*, edited by Thomas Albert Sebeok, 283–92. Cambridge, MA: M.I.T., 1960.
- Carson, D. A. "Three More Books on the Bible: A Critical Review." *Trinity Journal* 27 (2006) 1–62.
- Carson, D A., and Douglas J. Moo. *An Introduction to the New Testament*. 2nd ed. Grand Rapids: Zondervan, 2005.
- Cashdollar, Charles D. *The Transformation of Theology, 1830-1890: Positivism and Protestant Thought in Britain and America*. Princeton, N.J: Princeton University Press, 1989.
- Chan, Samuel W. K., and Mickey W. C. Chong. "Unsupervised Clustering for Nontextual Web Document Classification." *Decision Support Systems* 37 (2004) 377–96.

- Chaski, Carole E. "Empirical Evaluations of Language-Based Author Identification Techniques." *Forensic Linguistics* 8 (2001) 1–65.
- Cheng, Na, et al. "Author Gender Identification from Text." *Digital Investigation* 8 (2011) 78–88.
- Chirikjian, Gregory S. *Stochastic Models, Information Theory, and Lie Groups, Volume 2: Analytic Methods and Modern Applications*. Applied and Numerical Harmonic Analysis 2. Boston: Birkhäuser, 2012.
- Chomsky, Noam. *Syntactic Structures*. Janua Linguarum Series Minor 4. The Hague: Mouton, 1957.
- Chotlos, John W. "A Statistical and Comparative Analysis of Individual Written Language Samples." *Psychological Monographs: General and Applied* 56 (1944) 75–111.
- Clausen, Sten Erik. *Applied Correspondence Analysis: An Introduction*. Quantitative Applications in the Social Sciences 121. Thousand Oaks; London; New Dehli: Sage, 1998.
- Clayman, D. L. "Trends and Issues in Quantitative Stylistics." *Transactions of the American Philological Association (1974-2008)* 122 (1992) 385–90.
- Clogg, Clifford C. "Using Association Models in Sociological Research: Some Examples." *American Journal of Sociology* 88 (1982) 114–34.
- Clyde, James. *Greek Syntax, with a Rationale of the Constructions*. 6th ed. Edinburgh: Oliver and Boyd, 1881.
- Comte, Auguste. *Cours de philosophie positive*. Bachelier: Paris, 1830.
- Conzelmann, Hans, and Andreas Lindemann. *Interpreting the New Testament: An Introduction to the Principles and Methods of New Testament Exegesis*. Peabody, MA: Hendrickson, 1988.
- Cooper, R. A., and D. A. Pearsall. "The Gawain Poems: A Statistical Approach to the Question of Common Authorship." *The Review of English Studies* 39 (1988) 365–85.
- Cotterell, Peter, and Max Turner. *Linguistics & Biblical Interpretation*. Downers Grove: InterVarsity, 1989.
- Cover, Thomas M., et al. "Elements of Information Theory." *SIAM Review* 36 (1994) 509-10.
- Craig, Hugh. "Stylistic Analysis and Authorship Studies." In *A Companion to Digital Humanities*, edited by Susan Schreibman et al., 273–88. Blackwell Companion to Literature and Culture 26. Malden, MA: Blackwell, 2004.

- Crosby, A. *A Compendious Grammar of the Greek Language*. London; Gloucester: D. Bond; S. Harvard, 1870.
- Daelemans, Walter. "Explanation in Computational Stylometry." In *Computational Linguistics and Intelligent Text Processing*, edited by Alexander Gelbukh, 451–62. Lecture Notes in Computer Science 7817. Berlin; Heidelberg: Springer, 2013.
- Dai, Andrew M., and Amos J. Storkey. "The Supervised Hierarchical Dirichlet Process." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 37 (2015) 243-55.
- Damerau, Fred J. "The Use of Function Word Frequencies as Indicators of Style." *Computers and the Humanities* 9 (1975) 271–80.
- Davis, Boyd H., and Peyton R. Mason. *Locating Presence and Positions in Online Focus Group Text with Stance-Shift Analysis*. Hershey, PA: Information Science Reference/IGI Global, 2008.
- Davis, Boyd H., et al. "'It's a Wild Thing, Waiting to Get Me': Stance Analysis of African Americans with Diabetes." *The Diabetes Educator* 37 (2011) 409–18.
- Davis, James A. *The Logic of Causal Order. Quantitative Applications in the Social Sciences* 55. Thousand Oaks, CA: Sage, 1985.
- Davis, Philip W. *Modern Theories of Language*. Englewood Cliffs, N.J: Prentice-Hall, 1973.
- Davison, Michael E. "Paul v. Luke: A Computer Analysis of Some Differences." *Bulletin of the Association for Literary and Linguistic Computing* 12 (1984) 1–4.
- de Beaugrande, Robert. "'Register' in Discourse Studies: A Concept in Search of a Theory." In *Register Analysis: Theory and Practice*, edited by Mohsen Ghadessy, 7–25. London; New York: Pinter, 1993.
- de Morgan, Sophia E. *Memoir of Augustus de Morgan*. London: Longmans, Green, 1882.
- de Saussure, Ferdinand, et al. *Course in General Linguistics*. LaSalle, IL: Open Court, 1986.
- de Vel, O., et al. "Mining E-Mail Content for Author Identification Forensics." *ACM SIGMOD Record* 30 (2001) 55–64.
- de Wette, Wilhelm Martin Leberecht. *Kurze Erklärung der Briefe an die Colosser, an Philemon, an die Ephesier und Philipper*. Leipzig: Weidmann, 1843.
- . *Lehrbuch der historisch-kritischen Einleitung in die Bibel alten und neuen Testaments*. 4th ed. Berlin: Reimer, 1842.
- Deane, Philip. "Stylometrics Do Not Exclude the Seventh Letter." *Mind* 82 (1973) 113–17.

- Deissmann, Gustav Adolf. *Light from the Ancient East: The New Testament Illustrated by Recently Discovered Texts of the Graeco-Roman World*. Translated by Lionel Richard Mortimer Strachan. London: Hodder & Stoughton, 1927.
- Delbrück, Berthold. *Introduction to the Study of Language: A Critical Survey of the History and Methods of Comparative Philology of the Indo-European Languages*. Translated by Eva Channing. Leipzig: Breitkopf and Härtel, 1882.
- Delcourt, Christian. "About the Statistical Analysis of Co-occurrence." *Computers and the Humanities* 26 (1992) 21-29.
- . "Stylometry." *Revue Belge de Philologie et D'histoire* 80 (2002) 979–1002.
- DeMaris, Alfred. *Logit Modeling: Practical Applications*. Quantitative Applications in the Social Sciences 86. Newbury Park: Sage, 1992.
- Dewar, Roderick. "Information Theory Explanation of the Fluctuation Theorem, Maximum Entropy Production and Self-Organized Criticality in Non-Equilibrium Stationary States." *Journal of Physics A: Mathematical and General* 36 (2003) 631.
- Dewdney, Nigel , et al. "The Form is the Substance: Classification of Genres in Text." In *Proceedings of the Workshop on Human Language Technology and Knowledge Management*, 1–8. Toulouse, France: Association for Computational Linguistics, 2001.
- Dibelius, Martin. *Die Pastoralbriefe*. Lietzmann's Handbuch zum Neuen Testament 13. Tübingen: Mohr, 1955.
- Diederich, J., et al. "Authorship Attribution with Support Vector Machines." *Applied Intelligence* 19 (2003) 109–23.
- Diehl, Judith A. "What is a 'Gospel'? Recent Studies in the Gospel Genre." *Currents in Biblical Research* 9 (2011) 171–99.
- Dietterich, Thomas G. "Ensemble Methods in Machine Learning." In *Multiple Classifier Systems*, edited by J. Kittler and F. Roli, 1–15. Lecture Notes in Computer Science 1857. Berlin: Springer, 2000.
- . "An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization." *Machine Learning* 40 (2000) 139–57.
- Dinu, Liviu P., et al. "Pastiche Detection Based on Stopword Rankings: Exposing Impersonators of a Romanian Writer." In *Proceedings of the Workshop on Computational Approaches to Deception Detection*, 72–77. Avignon: Association for Computational Linguistics, 2012.

- Dittenberger, W. "Sprachliche Kriterien für die Chronologie der Platonischen Dialoge." *Hermes* 16 (1881) 321–45.
- Dixon, Peter, and David Mannion. "Goldsmith and the British Magazine." *Literary and Linguistic Computing* 13 (1998) 37–49.
- . "Goldsmith and the Public Ledger." *Language and Literature* 10 (2001) 307–23.
- Dixon, Robert M. W. *What is Language? A New Approach to Linguistic Description*. London: Longmans, 1965.
- Donelson, Lewis R. *Pseudepigraphy and Ethical Argument in the Pastoral Epistles*. Hermeneutische Untersuchungen zur Theologie 22. Tübingen: J.C.B. Mohr, 1986.
- Dormeyer, Detlev. *The New Testament Among the Writings of Antiquity*. Biblical Seminar Series 55. Sheffield: Sheffield Academic, 1998.
- Drake, Bryan. "Unanswered Questions in Computerized Literary Analysis." *Journal of Biblical Literature* 91 (1972) 241–42.
- Dreher, John J. "The Computer-Linguistic Detective of Authorship." *The Journal of Asian Studies* 29 (1970) 883–87.
- Drew, J., and H. Craig. "Did Dickens Write 'Temperate Temperance'? (An Attempt to Identify Authorship of an Anonymous Article in All the Year Round)." *Victorian Periodicals Review* 44 (2011) 267–90.
- Duff, Jeremy. "A Reconsideration of Pseudepigraphy in Early Christianity." D.Phil diss., Oxford University, 1998.
- Ebrahimpour, M., et al. "Automated Authorship Attribution Using Advanced Signal Classification Techniques." *PLOS* 8 (2013) 1–12.
- Edmundson, H. P. "Statistical Inference in Mathematical and Computational Linguistics." *International Journal of Computer & Information Sciences* 6 (1977) 95–129.
- Ehrman, Bart D. *Lost Scriptures: Books that Did Not Make It into the New Testament*. New York: Oxford University Press, 2003.
- Eichhorn, Johann Gottfried. *Einleitung in das Neue Testament*. Leipzig: Weidmann, 1812.
- Eissen, Sven Meyerzu, and Benno Stein. "Intrinsic Plagiarism Detection." In *Advances in Information Retrieval*, edited by Mounia Lalmas et al., 565–69. Lecture Notes in Computer Science 3936. Berlin; Heidelberg: Springer 2006.

- Ellegård, Alvar. *A Statistical Method for Determining Authorship: The Junius Letters, 1769-1772*. Acta Universitatis Gothoburgensis, Gothenburg Studies in English 13. Göteborg: Almqvist & Wiksell, 1962.
- Elliott, W. E. Y., and R. J. Valenza. "A Touchstone for the Bard." *Computers and the Humanities* 25 (1991) 199–209.
- Ellis, E. Earle. "The Pastorals and Paul." *The Expository Times* 104 (1992) 45–47.
- Emmert-Streib, Frank. *Information Theory and Statistical Learning*. Edited by Matthias Dehmer. Boston: Springer, 2009.
- Erasmus, Desiderius. *Annotationes, of, Aanteekeningen op 't Nieuwe Testament*. Amsterdam: Voor Ian Rievwertz, 1663.
- Erwin, Harry, and Michael Oakes. "Correspondence Analysis of the New Testament. Workshop on Language Resources and Evaluation for Religious Texts." Paper presented at the Language Resources and Evaluation for Religious Texts, Istanbul, Turkey, 2012.
- Evanson, Edward. *The Dissonance of the Four Generally Received Evangelists*. Ipswich: George Jermym, 1792.
- Faigley, Lester, and Paul Meyer. "Rhetorical Theory and Readers' Classifications of Text Types." *Text - Interdisciplinary Journal for the Study of Discourse* 3 (1983) 305–26.
- Fanning, Buist M. *Verbal Aspect in New Testament Greek*. Oxford: Oxford University Press, 1990.
- Fawcett, Robin P. *Cognitive Linguistics and Social Interaction: Towards an Integrated Model of a Systemic Functional Grammar and the Other Components of a Communicating Mind*. Exeter Linguistic Studies 3. Heidelberg: Groos, 1980.
- . *A Theory of Syntax for Systemic Functional Linguistics*. Amsterdam Studies in the Theory and History of Linguistic Science. Series IV, Current Issues in Linguistic Theory, 206. Amsterdam; Philadelphia: John Benjamins, 2000.
- Fee, Gordon. D. *New Testament Exegesis: A Handbook for Students and Pastors*. 3rd ed. Louisville: Westminster/John Knox, 2002.
- Ferrar, William Hugh. *A Comparative Grammar of Sanskrit, Greek, and Latin: In Two Volumes*. London; Dublin: Longmans, Green, Reader, and Dyer; William McGee, 1869.
- Finn, Aidan, et al. "Genre Classification and Domain Transfer for Information Filtering." In *Advances in Information Retrieval*, edited by Fabio Crestani et al., 353–62. Lecture Notes in Computer Science 2291. Berlin; Heidelberg: Springer, 2002.

- Finn, Patrick J. "Word Frequency, Information Theory, and Cloze Performance: A Transfer Feature Theory of Processing in Reading." *Reading Research Quarterly* 13 (1977) 508–37.
- Fiore, Benjamin. *The Function of Personal Example in the Socratic and Pastoral Epistles*. Analecta Biblica 105. Rome: Biblical Institute, 1986.
- Fiore, Benjamin, and Daniel J. Harrington. *The Pastoral Epistles: First Timothy, Second Timothy, Titus*. Sacra Pagina Series 12. Collegeville, MN: Liturgical, 2007.
- Fisher, Ronald Aylmer. "The Correlation between Relatives on the Supposition of Mendelian Inheritance." *Transactions of the Royal Society of Edinburgh* 52 (1918) 399–433.
- . *The Design of Experiments*. Edinburgh; London: Oliver and Boyd, 1935.
- . "On the "Probable Error" of a Coefficient of Correlation Deduced from a Small Sample." *Metron* 1 (1921) 3–32.
- . "On the Mathematical Foundations of Theoretical Statistics." *Philosophical Transactions of the Royal Society of London* 222 (1922) 309–68.
- . *Statistical Methods for Research Workers*. Biological Monographs and Manuals 4. Edinburgh; London: Oliver and Boyd, 1925.
- . "Uncertain Inference." *The Scientific Monthly* 43 (1936) 402–10.
- . "The Use of Multiple Measurements in Taxonomic Problems." *Annals of eugenics* 7 (1936) 179–88.
- Forbes, Dean A. "A Critique of Statistical Approaches to the Isaiah Authorship Problem." In *Bible et Informatique: Interprétation, Herméneutique, Compétence Informatique*, 531–45. Universität Tübingen: Champion, 1991.
- . "Statistical Research on the Bible." In *ABD* 6: 185–206. New York: Doubleday, 1992.
- Forsyth, Richard S., and David I. Holmes. "Feature-Finding for Text Classification." *Literary and Linguistic Computing* 11 (1996) 163–74.
- Forsyth, Richard S., et al. "Cicero, Sigonio, and Burrows: Investigating the Authenticity of the Consolatio." *Literary and Linguistic Computing* 14 (1999) 375–400.
- Foster, Donald W. "The Claremont Shakespeare Authorship Clinic: How Severe are The Problems?" *Computers and the Humanities* 32 (1998) 491–196.

- Fowl, Stephen E. *Engaging Scripture: A Model for Theological Interpretation*. Edited by Gareth Jones and Lewis Ayres. Challenges in Contemporary Theology 5. Blackwell: Malden, MA, 1998.
- . *The Theological Interpretation of Scripture: Classic and Contemporary Readings*. Malden, MA: Blackwell, 1997.
- Frantzeskou, Georgia, et al. "Source Code Authorship Analysis for Supporting the Cybercrime Investigation Process." In *Proceedings of the 1st International Conference on E-Business and Telecommunication Networks*, 85–92. Setúbal, Portugal: INSTICC, 2004.
- Frautschi, R. L. "Focal Style and the Problem of Attribution in Eight Prose Tales by Perrault." *Literary and Linguistic Computing* 2 (1987) 213–20.
- . "A Project for Author Discrimination in the Encyclopedie." *South Atlantic Bulletin* 32 (1967) 14–17.
- Frey, Jörg, ed. *Pseudepigraphie und Verfasserfiktion in frühchristlichen Briefen*. Wissenschaftliche Untersuchungen zum Neuen Testament 246. Tübingen: Mohr Siebeck.
- Friberg, Barbara, et al. *Analytical Greek New Testament: Greek Text Analysis*. Baker's Greek New Testament Library 1. Grand Rapids: Baker, 1981.
- Frow, John. *Genre*. London: Routledge, 2006.
- Fucks, Wilhelm. "On Mathematical Analysis of Style." *Biometrika* 39 (1952) 122–29.
- Gadamer, Hans-Georg. *Philosophical Hermeneutics*. Translated by David Edward Linge. Berkeley: University of California Press, 2008.
- Gani, J. "Characterizing an Author's Vocabulary." *South African Statistical Journal* 31 (1997) 1–11.
- . "Some Stochastic Models in Linguistic Analysis." *Advances in Applied Probability* 7 (1975) 232–34.
- Gardiner, Frederic. "The Language of the Epistle to the Hebrews as Bearing upon Its Authorship." *Journal of the Society of Biblical Literature and Exegesis* 7 (1887) 1–25.
- Genkin, Alexander, et al. "Large-Scale Bayesian Logistic Regression for Text Categorization." *Technometrics* 49 (2007) 291–304.
- Gerwig, G. W. "On the Decrease of Predication and of Sentence Weight in English Prose." *University Studies* 2 (1894) 17–44.
- Ghadessy, Mohsen, ed. *Thematic Development in English Texts*. London: Pinter, 1998.

- Gilula, Zvi, and J. Haberman Shelby. "The Analysis of Multivariate Contingency Tables by Restricted Canonical and Restricted Association Models." *Journal of the American Statistical Association* 83 (1988) 760–71.
- Girón, Javier, et al. "Bayesian Analysis of a Multinomial Sequence and Homogeneity of Literary Style." *The American Statistician* 59 (2005) 19–339.
- Gleason, Henry A. *Linguistics and English Grammar*. New York: Holt, Rinehart and Winston, 1965.
- Gliozzo, Alfio, and Carlo Strapparava. *Semantic Domains in Computational Linguistics*. Berlin; London: Springer, 2009.
- Gödel, Kurt. "Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme." *Monatshefte für Mathematik und Physik* 38 (1931) 173–98.
- Goetchius, Eugene Van Ness. *The Language of the New Testament*. New York: Scribner, 1965.
- Golan, Jonathan S. *The Linear Algebra a Beginning Graduate Student Ought to Know*. Dordrecht: Springer, 2012.
- Goodman, Leo A. "A General Model for the Analysis of Surveys." *American Journal of Sociology* 77 (1972) 1035–86.
- . "Some Useful Extensions of the Usual Correspondence Analysis Approach and the Usual Log-Linear Models Approach in the Analysis of Contingency Tables." *International Statistical Review / Revue Internationale de Statistique* 54 (1986) 243–70.
- Goodspeed, Edgar J. *The Key to Ephesians*. Chicago: University of Chicago Press, 1956.
- Grant, Tim. "Quantifying Evidence in Forensic Authorship Analysis." *International Journal of Speech Language and the Law* 14 (2007) 1–25.
- Gravetter, Frederick J., and Lori-Ann B. Forzano. *Research Methods for the Behavioral Sciences*. 4th ed. Belmont, CA: Cengage, 2011.
- Gray, David E. *Doing Research in the Real World*. Thousand Oaks, CA: Sage, 2004.
- Grayston, Kenneth, and G. Herdan. "Authorship of the Pastorals in the Light of Statistical Linguistics." *New Testament Studies* 6 (1959) 1–15.
- Green, Joel B. "Narrating the Gospel in 1 and 2 Peter." *Interpretation* 60 (2006) 262–77.
- . *Practicing Theological Interpretation: Engaging Biblical Texts for Faith and Formation*. Grand Rapids: Baker Academic, 2011.

- . “Scripture and Theology: Uniting the Two So Long Divided.” In *Between Two Horizons: Spanning New Testament Studies and Systematic Theology*, 23-43. Grand Rapids: Eerdmans, 2000.
- , ed. *Hearing the New Testament: Strategies for Interpretation*. 2nd ed. Grand Rapids: Eerdmans, 2010.
- Green, Joel B., and Max Turner, eds. *Between Two Horizons: Spanning New Testament Studies and Systematic Theology*. Grand Rapids: Eerdmans, 2000.
- Green, Paul E., and Donald S. Tull. *Research for Marketing Decisions*. 4th ed. Englewood Cliffs, N.J: Prentice-Hall, 1978.
- Greenacre, Michael J. *Correspondence Analysis in Practice*. 2nd ed. Boca Raton; London; New York: CRC, 2007.
- . *Theory and Applications of Correspondence Analysis*. London; Orlando: Academic, 1984.
- Greenacre, Michael J., and Jörg Blasius. *Multiple Correspondence Analysis and Related Methods*. Boca Raton, FL: Chapman & Hall/CRC, 2006.
- Greenberg, Joseph Harold. “The Measurement of Linguistic Diversity.” *Language* 32 (1956) 109–15.
- . “Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements.” In *Universals of Language*, edited by Joseph H. Greenberg, 73–113. London: MIT Press, 1963.
- Greenberg, Joseph Harold, et al., eds. *Universals of Human Language*. Stanford, CA: Stanford University Press, 1978.
- Greenspahn, Frederick E. “The Number and Distribution of Hapax Legomena in Biblical Hebrew.” *Vetus Testamentum* 30 (1980) 8–19.
- Greenwood, H. H. “Common Word Frequencies and Authorship in Luke’s Gospel and Acts.” *Literary and Linguistic Computing* 10 (1995) 183-87.
- . “Saint Paul Revisited: Word Clusters in Multidimensional Space.” *Literary and Linguistic Computing* 8 (1993) 211–19.
- . “St Paul Revisited: A Computational Result.” *Literary and Linguistic Computing* 7 (1992) 43–47.

- Grice, H. Paul. "Logic and Conversation." In *Studies in the Way of Words*, 22–40. Cambridge, MA: Harvard University Press, 1989.
- Gries, Stefan Thomas. *Multifactorial Analysis in Corpus Linguistics: A Study of Particle Placement*. New York: Continuum, 2003.
- Griffith, John G. "It's Greek to the Computer by A. Q. Morton; A. D. Winspear." *The Classical Review* 24 (1974) 162–63.
- Griffiths, Patrick, et al. *Language in Use: A Reader*. London; New York: Routledge, 2010.
- Grimm, Jakob. *Deutsche Grammatik*. Gottingen: Dieterich, 1840.
- . *Geschichte der Deutschen Sprache*. Leipzig: Weidmann, 1848.
- Guerin-Pace, France. "Textual Statistics. An Exploratory Tool for the Social Sciences." *Population: An English Selection* 10 (1998) 73–95.
- Guiraud, Pierre. *Les caractères statistiques du vocabulaire: essai de méthodologie*. Paris: Presses Universitaires de France, 1954.
- Guo, G., et al. *A Fast Document Classification Algorithm Based on Improved KNN*. Edited by J. S. Pan et al. ICICIC 2006: First International Conference on Innovative Computing, Information and Control, Vol 3, Proceedings Los Alamitos: IEEE Computer Society, 2006.
- Gurney, Penelope J., and Lyman W. Gurney. "Authorship Attribution of the *Scriptores Historiae Augustae*." *Literary & Linguistic Computing* 13 (1998) 119–31.
- . "Subsets and Homogeneity: Authorship Attribution in the *Scriptores Historiae Augustae*." *Literary & Linguistic Computing* 13 (1998) 133–40.
- Guthrie, Donald. "The Development of the Idea of Canonical Pseudepigrapha in New Testament Criticism." *Vox Evangelica* 1 (1962) 43–59.
- . *New Testament Introduction*. 3rd ed. Downers Grove: InterVarsity, 1970.
- . *The Pastoral Epistles: An Introduction and Commentary*. 2nd ed. Tyndale New Testament Commentaries 14. Leicester, UK; Grand Rapids: InterVarsity; Eerdmans, 1990.
- . *The Pastoral Epistles and the Mind of Paul*. London: Tyndale, 1956.
- Haberman, Shelby J. "The Analysis of Residuals in Cross-Classified Tables." *Biometrics* 29 (1973) 205–20.

- HaCohen-Kerner, Yaakov, et al. "Identifying Historical Period and Ethnic Origin of Documents Using Stylistic Feature Sets." In *Discovery Science*, edited by Ljupčo Todorovski et al., 102–13. Lecture Notes in Computer Science 4265. Berlin; Heidelberg: Springer 2006.
- . "Stylistic Feature Sets as Classifiers of Documents According to Their Historical Period and Ethnic Origin." *Applied Artificial Intelligence* 24 (2010) 847–62.
- Hájek, Petr, and Tomas Havránek. *Mechanizing Hypothesis Formation: Mathematical Foundations for a General Theory*. Berlin; New York: Springer, 1978.
- Hald, Anders. *A History of Mathematical Statistics from 1750 to 1930*. New York: Wiley, 1998.
- Halliday, M. A. K. "Categories of the Theory of Grammar." *Word* 17 (1961) 241–92.
- . *Computational and Quantitative Studies*. Edited by Jonathan J. Webster. Collected Works of M.A.K. Halliday 6. London; New York: Continuum, 2005.
- . *Explorations in the Functions of Language*. Explorations in Language Study 2. London: Edward Arnold, 1973.
- . *Halliday: System and Function in Language: Selected Papers*. Edited by G.R. Kress. London: Oxford University Press, 1976.
- . *Linguistic Studies of Text and Discourse*. Edited by Jonathan J. Webster. Collected works of M.A.K. Halliday 2. London: Continuum, 2002.
- . *On Grammar*. Edited by Jonathan J. Webster. Collected Works of M.A.K. Halliday 1. London; New York: Continuum, 2002.
- . *On Language and Linguistics*. Edited by Jonathan J. Webster. Collected Works of M.A.K. Halliday 3. New York: Continuum, 2003.
- Halliday, M. A. K., and Ruqaiya Hasan. *Cohesion in English*. English Language Series 9. London: Longman, 1976.
- . *Language, Context, and Text: Aspects of Language in a Social-Semiotic Perspective*. 2nd ed. Oxford: Oxford University Press, 1989.
- Halliday, M. A. K., and Christian M. I. M. Matthiessen. *An Introduction to Functional Grammar*. 3rd ed. London; New York: Arnold; Oxford University Press, 2004.
- Halliday, M. A. K., et al., eds. *New Developments in Systemic Linguistics*. Open Linguistics Series 2. London; New York: Pinter, 1987.
- Hardcastle, R. A. "CUSUM: A Credible Method for the Determination of Authorship?" *Science & Justice* 37 (1997) 129–38.

- Harding, M. *What Are They Saying About the Pastoral Epistles?* New York; Mahwah, NJ: Paulist, 2001.
- Harrington, D. J. *Interpreting the New Testament: A Practical Guide*. New Testament Message 1. Wilmington, DE: Liturgical, 1979.
- Harris, Zellig S. *Language and Information*. New York: Columbia University Press, 1988.
- . *Structural Linguistics*. Chicago; London: University of Chicago Press, 1951.
- Harrison, P. N. "Important Hypotheses Reconsidered." *The Expository Times* 67 (1955) 77–81.
- . *The Problem of the Pastoral Epistles*. London: Oxford University Press, 1921.
- Hartwell, Patrick M. "A Quantitative Approach to Thomas Hardy's Prose Style." *College Composition and Communication* 21 (1970) 150–56.
- Hartwig, Frederick, and Brian E. Dearing. *Exploratory Data Analysis*. Quantitative Applications in the Social Sciences 16. Newbury Park: Sage, 1979.
- Hasan, Ruqaiya. "Text in the Systemic-Functional Model." In *Current trends in textlinguistics*, edited by Wolfgang U. Dressler, 228–46. Research in Text Theory 2. Berlin; New York: W. de Gruyter, 1978.
- Hawkins, Douglas M., and David H. Olwell. *Cumulative Sum Charts and Charting for Quality Improvement*. Statistics for Engineering and Physical Science 1. New York: Springer, 1998.
- Herdan, Gustav. *The Advanced Theory of Language as Choice and Chance*. Kommunikation und Kybernetik in Einzeldarstellungen 4. Berlin: Springer, 1966.
- . *Language as Choice and Chance*. Groningen: P. Noordhoff, 1956.
- . *Type-Token Mathematics. A Textbook of Mathematical Linguistics*. Janua Linguarum Series Maior 4. Hague: Mouton, 1960.
- Hester, James D., and J. David Hester. *Rhetorics in the New Millennium: Promise and Fulfillment*. Studies in Antiquity & Christianity 14. New York: T & T Clark, 2010.
- Hilbert, David. "Die Grundlegung der elementaren Zahlenlehre." *Mathematische Annalen* 104 (1931) 485–94.
- Hilgenfeld, Adolf. *Historisch-kritische Einleitung in das Neue Testament*. Leipzig: Fues's (R. Reisland), 1875.

- Hilton, Michael L., and David I. Holmes. "An Assessment of Cumulative Sum Charts for Authorship Attribution." *Literary and Linguistic Computing* 8 (1993) 73–80.
- Hirsch, E. D. *Validity in Interpretation*. New Haven: Yale University Press, 1967.
- Hirschfeld, H. O. "A Connection between Correlation and Contingency." *Mathematical Proceedings of the Cambridge Philosophical Society* 31 (1935) 520–24.
- Hoey, Michael. *Lexical Priming: A New Theory of Words and Language*. London; New York: Routledge, 2005.
- Hoffmann, Michael. "Problems with Peirce's Concept of Abduction." *Foundations of Science* 4 (1999) 271–305.
- Holmes, David I. "The Analysis of Literary Style—A Review." *Journal of the Royal Statistical Society Series A (General)* 148 (1985) 328–41.
- . "Authorship Attribution." *Computers and the Humanities* 28 (1994) 87–116.
- . "The Evolution of Stylometry in Humanities Scholarship." *Literary and Linguistic Computing* 13 (1998) 111–17.
- . "A Stylometric Analysis of Mormon Scripture and Related Texts." *Journal of the Royal Statistical Society* 155 (1992) 91–120.
- Holmes, David I., and Richard S. Forsyth. "The Federalist Revisited: New Directions in Authorship Attribution." *Literary and Linguistic Computing* 10 (1995) 111–27.
- Holmes, David I., and Judit Kardos. "Who Was the Author? An Introduction to Stylometry." *Chance* 16 (2003) 5–8.
- Holmes, David I., et al. "A Widow and Her Soldier: Stylometry and the American Civil War." *Literary and Linguistic Computing* 16 (2001) 403–20.
- Holtzmann, Heinrich Julius. *Lehrbuch der historisch-kritischen Einleitung in das Neue Testament*. Freiburg: J.C.B. Mohr, 1885.
- Honoré, Antony. "Some Simple Measures of Richness of Vocabulary." *Association for Literary and Linguistic Computing Bulletin* 7 (1979) 172–77.
- Hoorn, Johan F., et al. "Neural Network Identification of Poets using Letter Sequences." *Literary and Linguistic Computing* 14 (1999) 311–38.
- Hoover, David L. "Another Perspective on Vocabulary Richness." *Computers and the Humanities* 37 (2003) 151–78.

- . “Multivariate Analysis and the Study of Style Variation.” *Literary and Linguistic Computing* 18 (2003) 341–60.
- Hoover, David L., and Shervin Hess. “An Exercise in Non-Ideal Authorship Attribution: The Mysterious Maria Ward.” *Literary and Linguistic Computing* 24 (2009) 467–89.
- Hopper, Vincent Foster. *Medieval Number Symbolism: Its Sources, Meaning, and Influence on Thought and Expression*. Mineola, NY: Dover, 2000.
- Horowitz, I. “Two Cultures of Science: The Limits of Positivism.” *ISSJ* 56 (2004) 429–37.
- Hotelling, Harold. “Analysis of a Complex of Statistical Variables into Principal Components.” *Journal of Educational Psychology* 24 (1933) 498–520.
- Householder, F.W. *Apollonius Dyscolus: The Syntax of Apollonius Dyscolus*. Studies in the History of the Language Sciences 23. Amsterdam: Benjamins, 1981.
- Howard, W.F. *The Gospel According to John*. The Interpreter’s Bible 8. Nashville: Abingdon, 1952.
- Hsu, Chich-Ming, and Ming-Syan Chen. “Subspace Clustering of High Dimensional Data with Noises [sic].” In *Advances in Knowledge Discovery and Data Mining: 8th Pacific-Asia Conference*, edited by Honghua Dai et al., 31–40. LNAI 3056. Berlin: Springer, 2004.
- Hudson, Richard A. *Sociolinguistics*. 2nd ed. Cambridge: Cambridge University Press, 1996.
- Humboldt, Wilhelm von. *Ueber die Kawi-Sprache auf der Insel Java: Nebst einer Einleitung über die Verschiedenheit des menschl. Sprachbaues & ihren Einfluss auf die geistige Entwicklung des Menschengeschlechts*. Berlin: Königlich-Preußische Akademie der Wissenschaften, 1836.
- Husson, François, et al. “Correspondence Analysis (CA).” In *Exploratory Multivariate Analysis by Example Using R*, 59–126. Boca Raton: Chapman & Hall; CRC, 2011.
- . *Exploratory Multivariate Analysis by Example Using R*. Boca Raton: Chapman & Hall; CRC, 2011.
- . “Principal Component Analysis (PCA).” In *Exploratory Multivariate Analysis by Example Using R*, 1–58. Boca Raton: Chapman & Hall; CRC, 2011.
- Iversen, Gudmund R. *Bayesian Statistical Inference*. Quantitative Applications in the Social Sciences 43. Beverly Hills: SAGE, 1984.
- Jamieson, Kathleen Hall, and Karlyn Kohrs Campbell. “Rhetorical Hybrids: Fusions of Generic Elements.” *Quarterly Journal of Speech* 68 (1982) 146–57.

- Jaynes, E. T. "How Does the Brain Do Plausible Reasoning?" In *Stanford University Microwave Laboratory*, 1–24. Stanford, CA: Stanford University Press, 1957.
- Jeanrond, Werner G. *Text and Interpretation as Categories of Theological Thinking*. New York: Crossroad, 1988.
- . *Theological Hermeneutics: Development and Significance*. New York: Crossroad, 1991.
- . "Theology in the Context of Pluralism and Postmodernity: David Tracy's Theological Method." In *Postmodernism, Literature and the Future of Theology*, edited by David Jasper, 143–63. Houndmills: Macmillan, 1993.
- Jefferson, Helen Genevieve. "Notes on the Authorship of Isaiah 65 and 66." *Journal of Biblical Literature* 68 (1949) 225–30.
- Jeffries, L., and D. McIntyre. *Stylistics*. Cambridge: Cambridge University Press, 2010.
- Jelf, William Edward. *A Grammar of the Greek Language: Volume II: Syntax*. 3rd ed. Oxford; London: John Henry and James Parker, 1861.
- Jeong, Hyeong Chul, et al. "Bootstrap Tests for Independence in Two-Way Ordinal Contingency Tables." *Computational Statistics & Data Analysis* 48 (2005) 623–31.
- Jin, Mingzhe, and Masakatsu Murakami. "Authors' Characteristic Writing Styles as Seen Through Their Use of Commas." *Behaviormetrika* 20 (1993) 63–76.
- Jockers, Matthew L. *Text Analysis with R for Students of Literature*. Heidelberg: Springer, 2010.
- Johnson, Luke Timothy. *The First and Second Letters to Timothy: A New Translation with Introduction and Commentary*. New York: Doubleday, 2001.
- Jordan, Michael I. "Graphical Models." *Statistical Science* 19 (2004) 140–55.
- Jülicher, Adolf. *An Introduction to the New Testament*. Translated by Janet Penrose Ward. London: Smith Elder & Company, 1904.
- Juola, Patrick. "Author Attribution." *Foundations and Trends in Information Retrieval* 1 (2008) 233–334.
- . "Authorship Attribution for Electronic Documents." In *Advances in Digital Forensics II*, edited by M. S. Olivier and S. Sheno, 119–30. International Federation for Information Processing 222. New York: Springer, 2006.
- . "Detecting Stylistic Deception." In *Proceedings of the Workshop on Computational Approaches to Deception Detection*, 91–96. Avignon, France: Association for Computational Linguistics, 2012.

- . “Future Trends in Authorship Attribution.” In *Advances in Digital Forensic III*, edited by P. Craiger and S. Sheno, 119–32. International Federation for Information Processing 242. New York: Springer, 2007.
- . “Large-Scale Experiments in Authorship Attribution.” *English Studies* 93 (2012) 275–83.
- Juola, Patrick, and R. Harald Baayen. “A Controlled-Corpus Experiment in Authorship Identification by Cross-Entropy.” *Literary and Linguistic Computing* 20 (2005) 59–67.
- Juola, Patrick, and Darren Vescovi. “Analyzing Stylometric Approaches to Author Obfuscation.” In *Advances in Digital Forensics VII*, edited by Gilbert Peterson and Sujeet Sheno, 115–25. IFIP Advances in Information and Communication Technology 361. Berlin: Springer, 2011.
- . “Empirical Evaluation of Authorship Obfuscation Using JGAAP.” In *Proceedings of the 3rd ACM Workshop on Artificial Intelligence and Security*, 14–18. Chicago: ACM, 2010.
- Juola, Patrick, et al. “A Prototype for Authorship Attribution Studies.” *Literary and Linguistic Computing* 21 (2006) 169–78.
- . “Towards Active Linguistic Authentication.” In *Advances in Digital Forensics IX*, edited by Gilbert Peterson and Sujeet Sheno, 385–98. IFIP Advances in Information and Communication Technology 410. New York: Springer, 2013.
- Kanaris, Ioannis, and Efstathios Stamatatos. “Learning to Recognize Webpage Genres.” *Information Processing and Management* 45 (2009) 499–512.
- Kantardzic, Mehmed. *Data Mining: Concepts, Models, Methods, and Algorithms*. 2nd ed. Hoboken: John Wiley & Sons, 2011.
- Karmeshu, ed. *Entropy Measures, Maximum Entropy Principle and Emerging Applications*. Studies in Fuzziness and Soft Computing 119. Berlin; Heidelberg: Springer, 2003.
- Kearney, Richard. “Ethics and the Postmodern Imagination.” *Thought: Fordham University Quarterly* 62 (1987) 39–58.
- Kennedy, George Alexander. *New Testament Interpretation Through Rhetorical Criticism*. Chapel Hill, NC: University of North Carolina Press, 1984.
- Kenny, Anthony. *The Computation of Style: An Introduction to Statistics for Students of Literature and Humanities*. Oxford: Pergamon, 1982.
- . *A Stylometric Study of the New Testament*. Oxford: Clarendon, 1986.

- Kestemont, Mike, et al. "Robust Rhymes? The Stability of Authorial Style in Medieval Narratives." *Journal of Quantitative Linguistics* 19 (2012) 54–76.
- Khmelev, Dmitri V., and Fiona J. Tweedie. "Using Markov Chains for Identification of Writers." *Literary and Linguistic Computing* 16 (2001) 299–307.
- Kim, Jae-On, and Charles W. Mueller. *Factor Analysis Statistical Methods and Practical Issues. Quantitative Applications in the Social Sciences* 14. Newbury Park: Sage, 1978.
- . *Introduction to Factor Analysis: What It Is and How to Do It. Quantitative Applications in the Social Sciences* 13. Newbury Park: Sage, 1978.
- Kjell, Bradley. "Authorship Attribution of Text Samples Using Neural Networks and Bayesian Classifiers." Paper presented at the 1994 IEEE International Conference on Systems, Man, and Cybernetics: Humans, Information and Technology, San Antonio, TX, Oct 2nd-5th, 1994.
- . "Authorship Determination Using Letter Pair Frequency Features with Neural Network Classifiers." *Literary and Linguistic Computing* 9 (1994) 119–24.
- Klauck, Hans-Josef. *The Religious Context of Early Christianity: A Guide to Graeco-Roman Religions*. Minneapolis: Fortress, 2003.
- Knight, George W. *The Pastoral Epistles: A Commentary on the Greek Text*. Grand Rapids: Eerdmans, 1992.
- Knoke, David, and Peter J. Burke. *Log-Linear Models. Quantitative Applications in the Social Sciences* 20. Newbury Park: SAGE, 1980.
- Knuth, Donald. E. *The Art of Computer Programming: Fundamental Algorithms*. Boston; London: Addison-Wesley, 2008.
- Kohavi, Ron. "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection." In *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI)*, 1137–45. San Francisco: Morgan Kaufmann, 1995.
- Kohavi, Ron, and George H. John. "Wrappers for Feature Subset Selection." *Artificial Intelligence* 97 (1997) 273–324.
- Koppel, Moshe, et al. "Authorship Attribution in the Wild." *Language Resources and Evaluation* 45 (2011) 83–94.
- . "Authorship Attribution with Thousands of Candidate Authors." In *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 659–60. Seattle, WA: ACM, 2006.

- . “Automatically Categorizing Written Texts by Author Gender.” *Literary and Linguistic Computing* 17 (2002) 401–12.
- . “Computational Methods in Authorship Attribution.” *Journal of the American Society for Information Science and Technology* 60 (2009) 9–26.
- . “Unsupervised Decomposition of a Document into Authorial Components.” In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, 1356–64. Portland, OR: Association for Computational Linguistics, 2011.
- Kornai, András. “Probabilistic Grammars and Languages.” *Journal of Logic, Language, and Information* 20 (2011) 317–28.
- Kripke, Saul. “Outline of a Theory of Truth.” *The Journal of Philosophy* 72 (1975) 690–716.
- Krippendorff, Klaus. *Information Theory. Quantitative Applications in the Social Sciences* 62. Newbury Park: Sage, 1986.
- Kucukyilmaz, Tayfun, et al. “Chat Mining for Gender Prediction.” In *Advances in Information Systems*, edited by Tatyana Yakhno and Erich J. Neuhold, 274–83. Lecture Notes in Computer Science 4243. Berlin; Heidelberg: Springer 2006.
- Kuhn, Thomas S. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press, 1962.
- . *The Structure of Scientific Revolutions*. International Encyclopedia of Unified Science 2. Chicago: University of Chicago Press, 1970.
- Kullback, Solomon. *Information Theory and Statistics*. New York: Dover, 1997.
- Kümmel, Werner Georg. *Introduction to the New Testament*. Translated by Howard Clark Kee. Nashville, TN: Abingdon, 1975.
- Kuta, Marcin, and Jacek Kitowski. “Optimisation of Character N-gram Profiles Method for Intrinsic Plagiarism Detection.” In *Artificial Intelligence and Soft Computing*, edited by Leszek Rutkowski et al., 500–11. Lecture Notes in Computer Science 8468. Springer, 2014.
- Kwong, Ivan Shing Chung. *The Word Order of the Gospel of Luke: Its Foreground Messages*. Library of New Testament Studies 12. London; New York: T & T Clark, 2005.
- Labbé, Dominique. “Experiments on Authorship Attribution by Intertextual Distance in English.” *Journal of Quantitative Linguistics* 14 (2007) 33–80.

- Lamb, Sydney. M., and Leonard. E. Newell. *An Outline of Stratificational Grammar*. Washington, DC: Georgetown University Press, 1966.
- Lambers, Maarten, and Cor J. Veenman. "Forensic Authorship Attribution Using Compression Distances to Prototypes." In *Computational Forensics*, edited by Zeno J. M. H. Geradts et al., 13–24. Lecture Notes in Computer Science 5718. Berlin; Heidelberg: Springer 2009.
- Lambrecht, Jan. "Rhetorical Criticism and the New Testament." *Bijdragen* 50 (1989) 239–53.
- Lana, Maurizio. "Xenophon's Athenaion Politeia: A Study by Correspondence Analysis." *Literary and Linguistic Computing* 7 (1992) 17–26.
- Larsen, Wayne A., et al. "Who Wrote the Book of Mormon? An Analysis of Wordprints." *BYU Studies* 20 (1980) 225-51.
- Layton, Robert, et al. "Automated Unsupervised Authorship Analysis Using Evidence Accumulation Clustering." *Natural Language Engineering* 19 (2013) 95–120.
- Le Roux, Brigitte, and Henry Rouanet. *Multiple Correspondence Analysis. Quantitative Applications in the Social Sciences* 163. Thousand Oaks: Sage, 2010.
- Ledger, Gerard R. "An Exploration of Differences in the Pauline Epistles Using Multivariate Statistical Analysis." *Literary and Linguistic Computing* 10 (1995) 85–97.
- . "A New Approach to Stylometry." *Bulletin of the Association for Literary and Linguistic Computing* 13 (1985) 67–72.
- Ledger, Gerard R., and Thomas Merriam. "Shakespeare, Fletcher, and the Two Noble Kinsmen." *Literary and Linguistic Computing* 9 (1994) 235–48.
- Lehmann, Winfried Philipp. *Descriptive Linguistics: An Introduction*. New York: Random House, 1976.
- Libby, James A. "Locating Computational Methods within the Theological Spectrum: A Reflection." Forthcoming.
- Liebetrau, Albert M. *Measures of Association. Quantitative Applications in the Social Sciences* 32. Newbury Park: Sage, 1983.
- Lincoln, Andrew T. *Ephesians*. Word Biblical Commentary 42. Nashville, TN: Thomas Nelson, 1990.
- Linmans, A. J. M. "Correspondence Analysis of the Synoptic Gospels." *Literary and Linguistic Computing* 13 (1998) 1–13.
- Lipschutz, Seymour, et al. *Linear Algebra*. New York: McGraw-Hill, 2002.

- Loehlin, John C. "Component Analysis Versus Common Factor Analysis: A Case of Disputed Authorship." *Multivariate Behavioral Research* 25 (1990) 29–31.
- Long, J. Scott. *Confirmatory Factor Analysis: A Preface to LISREL*. Quantitative Applications in the Social Sciences 33. Beverly Hills: Sage, 1983.
- . *Covariance Structure Models: An Introduction to LISREL*. Quantitative Applications in the Social Sciences 34. Newbury Park: Sage, 1983.
- Louw, Johannes P. *Semantics of New Testament Greek*. Semeia Studies 11. Atlanta, GA: Scholars, 1982.
- Louw, Johannes P., and Eugene A. Nida, eds. *Greek-English Lexicon of the New Testament: Based on Semantic Domains, Volume 1*. New York: United Bible Societies, 1989.
- , eds. *Greek-English Lexicon of the New Testament: Based on Semantic Domains, Volume 2*. New York: United Bible Societies, 1989.
- Love, Harold. *Attributing Authorship: An Introduction*. Cambridge; New York: Cambridge University Press, 2002.
- Lowe, David, and Robert Matthews. "Shakespeare vs. Fletcher: A Stylometric Analysis by Radial Basis Functions." *Computers and the Humanities* 29 (1995) 449–61.
- Lukaszewski, Albert L. *The Lexham Syntactic Greek New Testament*. Bellingham, WA: Logos Research Systems, 2006.
- Lutosławski, Wincenty. "On Stylometry." *The Classical Review* 11 (1897) 284–86.
- . *The Origin and Growth of Plato's Logic: With an Account of Plato's Style and of the Chronology of His Writings*. London: Longmans & Green, 1897.
- . "Principes de stylométrie appliqués à la chronologie des oeuvres de Platon." *Revue des études grecques* 11 (1898) 61–81.
- Luyckx, Kim. *Scalability Issues in Authorship Attribution*. Brussels, Belgium: University Press Antwerp, 2010.
- Luyckx, Kim, and Walter Daelemans. "Authorship Attribution and Verification with Many Authors and Limited Data." In *Proceedings of the 22nd International Conference on Computational Linguistics*, 513–20. Manchester, UK: Association for Computational Linguistics, 2008.
- . "The Effect of Author Set Size and Data Size in Authorship Attribution." *Literary and Linguistic Computing* 26 (2011) 35–55.

- . “Shallow Text Analysis and Machine Learning for Authorship Attribution.” In *Proceedings of the 15th Meeting of Computational Linguistics in the Netherlands*, 149–60. Utrecht, Netherlands: LOT, 2005.
- Luz, Ulrich. *Studies in Matthew*. Grand Rapids: Eerdmans, 2005.
- Lyons, John. *Introduction to Theoretical Linguistics*. Cambridge: Cambridge University Press, 1969.
- Mack, Burton L. *Rhetoric and the New Testament*. Guides to Biblical Scholarship / New Testament Series 5. Minneapolis: Augsburg Fortress, 1990.
- Mackonis, Adolfas. “Inference to the Best Explanation, Coherence and Other Explanatory Virtues.” *Synthese* 190 (2013) 975–95.
- Maddalena, Giovanni. “Abduction and Metaphysical Realism.” *Semiotica* 153 (2005) 243–59.
- Madigan, David, et al. “Author Identification on the Large Scale.” Paper presented at the Proceedings of the 2005 Joint Annual Meeting of the Interface and Classification Society of North America, St. Louis, MO, 2005.
- Magnani, Lorenzo. “An Abductive Theory of Scientific Reasoning.” *Semiotica* 153 (2005) 261–86.
- Maloney, Gilles, et al. “La longueur des mots comme mesure d'homogeneite chez Hippocrate.” *Phoenix* 31 (1977) 95–122.
- Malyutov, M. B. “Authorship Attribution of Texts: A Review.” In *General Theory of Information Transfer and Combinatorics*, edited by Rudolf Ahlswede et al., 362–80. Lecture Notes in Computer Science 4123. Berlin: Springer, 2006.
- Mandolfo, Carleen. “Dialogic Form Criticism: An Intertextual Reading of Lamentations and Psalms of Lament.” In *Bakhtin and Genre Theory in Biblical Studies*, edited by Roland Boer, 69-90. Atlanta: Society of Biblical Literature, 2007.
- Markov, A. A. *Ischislenie veroiatnostei*. Moscow: Gos. izd-vo, 1924.
- . “Пример статистического исследования над текстом 'Евгения Онегина', иллюстрирующий связь испытаний в цепь.” *Известия Имп. Акад. наук, серия VI* 10 (1913) 153–62.
- Marriott, Ian. “The Authorship of the *Historia Augusta*: Two Computer Studies.” *Journal of Roman Studies* 69 (1979) 65–77.

- Marsden, John, et al. "Language Individuation and Marker Words: Shakespeare and His Maxwell's Demon." In *Plos One*, 1-12. San Francisco: Public Library of Science, 2013.
- Marshall, I. Howard. *New Testament Theology: Many Witnesses, One Gospel*. Downers Grove: InterVarsity, 2004.
- . "'Sometimes Only Orthodox'—Is There More to the Pastoral Epistles?" *Epworth Review* 20 (1993) 12–24.
- Martín-Asensio, Gustavo. *Transitivity-Based Foreground in the Acts of the Apostles: A Functional-Grammatical Approach to the Lukan Perspective*. Journal for the Study of the New Testament. 8. Sheffield: Sheffield Academic, 2000.
- Martin, James R. *English Text: System and Structure*. Philadelphia: John Benjamins, 1992.
- Martin, James R., and David Rose. *Genre Relations: Mapping Culture*. Oakville, CT: Equinox, 2008.
- Martin, James R., and Peter R. R. White. *The Language of Evaluation*. Basingstoke; New York: Palgrave Macmillan, 2005.
- Martindale, Colin, and Dean McKenzie. "On the Utility of Content Analysis in Author Attribution: The Federalist." *Computers and the Humanities* 29 (1995) 259–70.
- Marton, Yuval, et al. "On Compression-Based Text Classification." In *Advances in Information Retrieval Research*, edited by D.E. Losada and J.M. Fernandez-Luna, 300–14. LNCS 3408. Berlin; Heidelberg: Springer, 2005.
- Mascol, Conrad. "Curves of Pauline and Pseudo-Pauline Style I." *Unitarian Review* 30 (1888) 452–60.
- . "Curves of Pauline and Pseudo-Pauline Style II." *Unitarian Review* 30 (1888) 539–46.
- Mason, Jane E., et al. "Classifying Web Pages by Genre: An N-Gram Approach." Paper presented at the Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology, Milan, Italy, 15-18 Sept 2009, 2009.
- Mastro Paolo, Antonio, et al. "Legal Documents Categorization by Compression." In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Law*, 92–100. Rome: ACM, 2013.
- Mathewson, David. *Verbal Aspect in the Book of Revelation: The Function of Greek Verb Tenses in John's Apocalypse*. Linguistic Biblical Studies 4. Leiden; Boston: Brill, 2010.

- Matlack, Cynthia S., and William F. Matlack. "A Statistical Approach to Problems of Attribution: A Letter of Advice to a Young Poet." *College English* 29 (1968) 627–32.
- Matthews, Robert A. J., and Thomas V. N. Merriam. "Neural Computation in Stylometry I: An Application to the Works of Shakespeare and Fletcher." *Literary and Linguistic Computing* 8 (1993) 203–09.
- Matthiessen, C.M.I.M. "Register in the Round: Diversity in a Unified Theory of Register Analysis." In *Register Analysis: Theory and Practice*, edited by Mohsen Ghadessy, 221–292. London: Pinter, 1993.
- McColly, William B. "Style and Structure in the Middle English Poem 'Cleanness'." *Computers and the Humanities* 21 (1987) 169–76.
- McGann, Jerome. "Marking Texts of Many Dimensions." In *A Companion to Digital Humanities*, edited by Susan Schreibman et al., 198–217. Blackwell Companions to Literature and Culture 26. Malden, MA: Blackwell, 2004.
- McKaughan, Daniel J. "From Ugly Duckling to Swan: C. S. Peirce, Abduction, and the Pursuit of Scientific Theories." *Transactions of the Charles S. Peirce Society: A Quarterly Journal in American Philosophy* 44 (2008) 446–68.
- McKnight, Scot. *Introducing New Testament Interpretation*. Guides to New Testament Exegesis 1. Grand Rapids: Baker, 1990.
- McLaren, Brian D. *The Story We Find Ourselves In: Further Adventures of a New Kind of Christian*. San Francisco: Jossey-Bass, 2003.
- McLean, B. Hudson, ed. *Origins and Method: Towards a New Understanding of Judaism and Christianity: Essays in Honour of John C. Hurd*. Journal for the Study of the New Testament 86. Sheffield: JSOT, 1993.
- McLeod, Stephen K. "Knowledge of Necessity: Logical Positivism and Kripkean Essentialism." *Philosophy* 83 (2008) 179–91.
- McQueen, David V. "The Nature of Causality: Beyond Traditional Evidence." In *Global Handbook on Noncommunicable Diseases and Health Promotion*, edited by David V. McQueen, 37–49. New York: Springer, 2013.
- Mealand, David L. "Computers in New Testament Research." *Journal for the Study of the New Testament* 10 (1988) 97–115.
- . "The Extent of the Pauline Corpus: A Multivariate Approach." *Journal for the Study of the New Testament* 18 (1995) 61–92.

- . “Hellenistic Greek and the New Testament: A Stylometric Perspective.” *Journal for the Study of the New Testament* 34 (2012) 323–45.
- . “Is There Stylometric Evidence for Q?” *New Testament Studies* 57 (2011) 483–507.
- . “Measuring Genre Differences in Mark with Correspondence Analysis.” *Literary & Linguistic Computing* 12 (1997) 227–45.
- . “Positional Stylometry Reassessed: Testing a Seven Epistle Theory of Pauline Authorship.” *New Testament Studies* 35 (1989) 266–86.
- . “Style, Genre, and Authorship in Acts, the Septuagint, and Hellenistic Historians.” *Literary & Linguistic Computing* 14 (1999) 479–505.
- Meeks, Wayne A. “Forward.” In *The New Testament in its Literary Environment*, 7–8. Library of Early Christianity 8. Philadelphia: Westminster, 1987.
- Mendenhall, T. C. “The Characteristic Curves of Composition.” *Science* 9 (1887) 237–49.
- . “A Mechanical Solution of a Literary Problem.” *Popular Science Monthly* LX (1901) 97–105.
- Merriam, Thomas V. N. “The Authorship of Sir Thomas More.” *Bulletin of the Association for Literary and Linguistic Computing* 10 (1982) 1–7.
- . “Edward III.” *Literary and Linguistic Computing* 15 (2000) 157–86.
- . “An Experiment with the Federalist Papers.” *Computers and the Humanities* 23 (1989) 251–54.
- . “Heterogeneous Authorship in Early Shakespeare and the Problem of Henry V.” *Literary & Linguistic Computing* 13 (1998) 15–28.
- . “Invalidation Reappraised.” *Computers and the Humanities* 30 (1996) 417–31.
- . “An Investigation of Morton’s Method: A Reply.” *Computers and the Humanities* 21 (1987) 57–58.
- . “Letter Frequency as a Discriminator of Authors.” *Notes and Queries* 41 (1994) 467–69.
- . “Marlowe and Nashe in Dido Queen of Carthage.” *Notes and Queries* 47 (2000) 425–28.
- . “Marlowe’s Hand in Edward III.” *Literary and Linguistic Computing* 8 (1993) 59–72.
- . “Tamburlaine Stalks in ‘Henry VI.’” *Computers and the Humanities* 30 (1996) 267–80.

- . “An Unwarranted Assumption (On the Authorship of 'Sir Thomas More').” *Notes and Queries* 47 (2000) 438–41.
- Merriam, Thomas V. N., and Robert A. J. Matthews. “Neural Computation in Stylometry II: An Application to the Works of Shakespeare and Marlowe.” *Literary and Linguistic Computing* 9 (1994) 1–6.
- Metzger, Bruce. M. *The Canon of the New Testament: Its Origin, Development, and Significance*. Oxford: Clarendon, 1997.
- Meulman, J., and W. J. Heiser. “SPSS Categories 21.0.” Online PDF for SPSS Version 21.0: SPSS Incorporated, 2012.
- Meyers, Lawrence S., et al. *Applied Multivariate Research: Design and Interpretation*. Thousand Oaks, CA: Sage, 2013.
- Meyn, Sean P., and Richard L. Tweedie. *Markov Chains and Stochastic Stability*. Communications and Control Engineering 16. London; New York: Springer, 1993.
- Michaelis, Wilhelm “Pastoralbriefe und Wortstatistik.” *Zeitschrift für die neutestamentliche Wissenschaft und die Kunde der älteren Kirche* 28 (1929) 69–76.
- Michaelson, S., and A. Q. Morton. “The New Stylometry: A One-Word Test of Authorship for Greek Writers.” *The Classical Quarterly* 22 (1972) 89–102.
- Michell, Joel. “Pragmatism, Positivism and the Quantitative Imperative.” *Theory & Psychology* 13 (2003) 45–52.
- Mitchell, Margaret M. “P^Tebt 703 and the Genre of 1 Timothy: The Curious Career of a Ptolemaic Papyrus in Pauline Scholarship.” *Novum Testamentum* 44 (2002) 344–70.
- Moerk, Ernst L. “An Objective, Statistical Description of Style.” *Linguistics* 108 (1973) 50–58.
- Moffatt, James. *An Introduction to the Literature of the New Testament*. New York: Charles Scribner’s Sons, 1911.
- Moles, Abraham A. *Information Theory and Esthetic Perception*. Translated by Joel E. Cohen. Urbana; London: University of Illinois Press, 1966.
- Mollet, Eugène, et al. “Choosing the Best Tools for Comparative Analyses of Texts.” *International Journal of Corpus Linguistics* 15 (2010) 429–73.
- Montgomery Hitchcock, F. R. “Latinity of the Pastorals.” *The Expository Times* 39 (1928) 347–52.

- Montgomery, John Warwick. "Theologian's Craft: A Discussion of Theory Formation and Theory Testing in Theology." *Concordia Theological Monthly* 37 (1966) 67–98.
- Mooney, Christopher Z., and Robert D. Duval. *Bootstrapping: A Nonparametric Approach to Statistical Inference*. Quantitative Applications in the Social Sciences 95. Newbury Park: Sage, 1993.
- Moore, Stephen D. *The Bible in Theory: Critical and Postcritical Essays*. Resources for Biblical Study 57. Atlanta: Society of Biblical Literature, 2010.
- Moore, Stephen D., and Yvonne Sherwood. "Biblical Studies 'After' Theory: Onwards Towards the Past. Part 2, The Secret Vices of the Biblical God." *Biblical Interpretation* 18 (2010) 87–113.
- . "Biblical Studies 'After' Theory: Onwards Towards the Past. Part 3, Theory in the First and Second Waves." *Biblical Interpretation* 18 (2010) 191–225.
- Morgan, Brian. "Poststructuralism and Applied Linguistics." In *International Handbook of English Language Teaching*, edited by Jim Cummins and Chris Davison, 1033–52. Springer International Handbooks of Education 15. New York: Springer, 2007.
- Moritz, Robert Édouard. "On the Variation and Functional Relation of Certain Sentence-Constants in Standard Literature." *University Studies* 3 (1903) 229–53.
- Morton, Andrew Queen. "The Authorship of Greek Prose." *Journal of the Royal Statistical Society* 128 (1965) 169–233.
- . *Literary Detection: How to Prove Authorship and Fraud in Literature and Documents*. New York: Scribner, 1978.
- . "Once. A Test of Authorship Based on Words Which are Not Repeated in the Sample." *Literary and Linguistic Computing* 1 (1986) 1–8.
- Morton, Andrew Queen, and James McLeman. *Paul, the Man and the Myth; A Study in the Authorship of Greek Prose*. New York: Harper & Row, 1966.
- Morton, Andrew Queen, and Alban Dewes Winspear. *It's Greek to the Computer*. Edited by Alban Dewes Winspear. Montreal: Harvest House, 1971.
- Mosteller, Frederick. *Applied Bayesian and Classical Inference: The Case of the Federalist Papers*. 2nd ed. Springer Series in Statistics 14. New York: Springer, 1984.
- . "A Statistical Study of the Writing Styles of the Authors of 'The Federalist' Papers." *Proceedings of the American Philosophical Society* 131 (1987) 132–40.

- . “Who Wrote the Disputed Federalist Papers, Hamilton or Madison?” In *Pleasure of Statistics: The Autobiography of Frederick Mosteller*, edited by Stephen Fienberg, E. et al., 47–67. New York: Springer, 2010.
- Mosteller, Frederick, and David L. Wallace. *Inference and Disputed Authorship: The Federalist*. Reading, MA: Addison-Wesley, 1964.
- . *Inference and Disputed Authorship: The Federalist*. Stanford: Center for the Study of Language and Information, 2007.
- . “Inference in an Authorship Problem.” *Journal of the American Statistical Association* 58 (1963) 275–309.
- Mounce, William D. *Basics of Biblical Greek Grammar*. Grand Rapids: Zondervan, 2009.
- Mukherjee, Arjun, and Bing Liu. “Improving Gender Classification of Blog Authors.” In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, 207–17. Cambridge, MA: Association for Computational Linguistics, 2010.
- Mullins, Phil. “Peirce’s Abduction and Polanyi’s Tacit Knowing.” *The Journal of Speculative Philosophy* 16 (2002) 198–224.
- Murphy-O’Connor, Jerome. “2 Timothy Contrasted with 1 Timothy and Titus.” *Revue Biblique* 98 (1991) 403–18.
- Narayanan, Arvind, et al. “On the Feasibility of Internet-Scale Author Identification.” In *2012 IEEE Symposium on Security and Privacy (Sp)*, 300–14. San Francisco, CA: IEEE, 2012.
- Nelder, J. A. “Log Linear Models for Contingency Tables: A Generalization of Classical Least Squares.” *Journal of the Royal Statistical Society* 23 (1974) 323–29.
- Nelder, J. A., and R. W. M. Wedderburn. “Generalized Linear Models.” *Journal of the Royal Statistical Society* 135 (1972) 370–84.
- Neumann, Kenneth J. *The Authenticity of the Pauline Epistles in the Light of Stylostatistical Analysis*. Society of Biblical Literature Dissertation Series 120. Atlanta, GA: Scholars, 1990.
- . “Major Variations in Pauline and Other Epistles in Light of Genre and the Pauline Letter Form.” In *Origins and Method: Towards A New Understanding of Judaism and Christianity: Essays in Honour of John C. Hurd*, edited by Bradley H. McLean, 199–209. Journal for the Study of the New Testament Supplement Series 86. Sheffield: Sheffield Academic, 1993.
- Neurath, Otto. “Physicalism: The Philosophy of the Viennese Circle.” *The Monist* 41 (1931) 618–23.

- Nida, Eugene Albert, and Johannes P. Louw. *Lexical Semantics of the Greek New Testament: A Supplement to the Greek-English Lexicon of the New Testament Based on Semantic Domains*. Resources for Biblical Study 25. Atlanta, GA: Scholars, 1992.
- Nini, Andrea, and Tim Grant. "Bridging the Gap between Stylistic and Cognitive Approaches to Authorship Analysis Using Systemic Functional Linguistics and Multidimensional Analysis." *International Journal of Speech Language and the Law* 20 (2013) 173–202.
- Nollen, J. S. "Qualities of Style as a Test of Authorship: A Criticism of Wolff's *Zwei Jugendspiele von Heinrich von Kleist*." *PMLA: Publications of the Modern Language Association of America* 14, supplement (1899) lxix–lxx.
- O'Brien, P. T. *The Letter to the Ephesians*. Grand Rapids: Eerdmans, 1999.
- O'Donnell, Bernard. "Stephen Crane's the O'Ruddy: A Problem in Authorship Discrimination." In *The Computer & Literary Style: Introductory Essays and Studies*, edited by Jacob Leed, 107–15. Kent Studies in English 2. Kent, OH: Kent State University Press, 1966.
- O'Donnell, Matthew Brook. *Corpus Linguistics and the Greek of the New Testament*. New Testament Monographs 6. Sheffield: Sheffield Phoenix, 2005.
- . "Linguistic Fingerprints or Style by Numbers? The Use of Statistics in the Discussion of Authorship of New Testament Documents." In *Linguistics and the New Testament: Critical Junctures*, edited by Stanley E. Porter and D. A. Carson, 206–26. Journal for the Study of the New Testament Supplement Series 168. Sheffield: Sheffield Academic, 1999.
- . "The Use of Annotated Corpora for New Testament Discourse Analysis: A Survey of Current Practice and Future Prospects." In *Discourse Analysis and the New Testament: Approaches and Results*, edited by Stanley E. Porter and Jeffrey T. Reed, 71–116. Journal for the Study of the New Testament Supplement Series 170. Sheffield: Sheffield Academic, 1999.
- O'Donnell, Matthew Brook, et al. "OpenText.org and the Problems and Prospects of Working with Ancient Discourse." In *A Rainbow of Corpora: Corpus Linguistics and the Languages of the World*, edited by A.P. Wilson et al., 109–21. Munich: Lincom, 2003.
- O'Keefe, Richard A. "Critical Remarks on Houk's 'Statistical Analysis of Genesis Sources'." *Journal for the Study of the Old Testament* 29 (2005) 409–37.
- Oakman, Robert L. *Computer Methods for Literary Research*. Columbia, S.C.: University of South Carolina Press, 1980.
- . *Computer Methods for Literary Research*. 2nd ed. Columbia, S.C.: University of South Carolina Press, 1984.

- Oberreuter, Gabriel, and Juan D. Velásquez. "Text Mining Applied to Plagiarism Detection: The Use of Words for Detecting Deviations in the Writing Style." *Expert Systems with Applications* 40 (2013) 3756–63.
- Ochs, Peter. "Philosophic Warrants for Scriptural Reasoning." *Modern Theology* 22 (2006) 465–82.
- Oliveira, W., et al. "Comparing Compression Models for Authorship Attribution." *Forensic Science International* 228 (2013) 100–04.
- Paavola, Sami. "Abduction as a Logic and Methodology of Discovery: The Importance of Strategies." *Foundations of Science* 9 (2004) 267–83.
- Paltridge, Brian. "Genre Analysis and the Identification of Textual Boundaries." *Applied Linguistics* 15 (1994) 288–99.
- Paolillo, John C., et al. "Genre Emergence in Amateur Flash." In *Genres on the Web Computational Models and Empirical Studies*, edited by Alexander Mehler et al., 277–302. Dordrecht; New York: Springer, 2010.
- Park, Seong-Bae, and Byoung-Tak Zhang. "Co-trained Support Vector Machines for Large Scale Unstructured Document Classification Using Unlabeled Data and Syntactic Information." *Information Processing & Management* 40 (2004) 421–39.
- . "Large Scale Unstructured Document Classification Using Unlabeled Data and Syntactic Information." In *Advances in Knowledge Discovery and Data Mining*, edited by K. Y. Whang et al., 88–99. Lecture Notes in Artificial Intelligence 2637. Berlin: Springer, 2003.
- Parry, Reginald St John. *The Pastoral Epistles with Introduction, Text and Commentary*. Cambridge: Cambridge University Press, 1920.
- Parunak, H. van Dyke. "Computers and Biblical Studies." In *ABD* 1: 1112–24. New York: Doubleday, 1992.
- Pawlowski, Adam, and Artur Pacewicz. "Wincenty Lutoslawski: Philosophe, helléniste ou fondateur sous-estimé de la stylométrie?" *Historiographia Linguistica* 31 (2004) 423–47.
- Pearl, Judea. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Morgan Kaufmann, 1988.
- Pearson, Brook W. R., and Stanley E. Porter. "The Genres of the New Testament." In *Handbook to Exegesis of the New Testament*, edited by Stanley E. Porter, 131–66. New Testament Tools and Studies 25. Boston: Brill Academic, 2002.

- Peirce, Charles Sanders. *Collected Papers of Charles Sanders Peirce*. Edited by Arthur Walter Burks. Volume 7-8. Cambridge, MA: Harvard University Press, 1958.
- . “Deduction, Induction, and Hypothesis.” *Popular Science Monthly* 13 (1878) 470–82.
- . “On the Logic of Drawing History from Ancient Documents, Especially from Testimonies.” In *The Essential Peirce: Selected Philosophical Writings, (1893–1913)*, edited by The Peirce Edition Project, 75–114. Bloomington: Indiana University Press, 1901.
- . “On the Natural Classification of Arguments.” *Proceedings of the American Academy of Arts and Sciences* 7 (1867) 261–87.
- Peng, Roger D., and Nicolas W. Hengartner. “Quantitative Analysis of Literary Styles.” *The American Statistician* 56 (2002) 175–85.
- Penick, Daniel A. “Paul’s Epistles Compared with One Another and with the Epistle to the Hebrews.” *The American Journal of Philology* 42 (1921) 58–72.
- Pereira, Fernando. “Formal Grammar and Information Theory: Together Again?” *Philosophical Transactions: Mathematical, Physical and Engineering Sciences* 358 (2000) 1239–53.
- Petruszewycz, Micheline. “Chaînes de Markov et statistiques linguistiques.” *Mots* 7 (1983) 85–95.
- . *Les chaînes de Markov dans le domaine linguistique*. Geneva: Slatkine, 1981.
- Pierce, J.R. *An Introduction to Information Theory: Symbols, Signals and Noise*. New York: Dover, 2012.
- Pike, Kenneth L. *Language in Relation to a Unified Theory of the Structure of Human Behavior*. Glendale, CA: Summer Institute of Linguistics, 1954.
- . *Language in Relation to a Unified Theory of the Structure of Human Behavior*. 2nd ed. Paris: Moulton, 1967.
- . *Linguistic Concepts: An Introduction to Tagmemics*. Lincoln: University of Nebraska Press, 1982.
- Pitts, Andrew W. “Greek Case in the Hellenistic and Byzantine Grammarians.” In *The Language of the New Testament: Context, History and Development*, edited by Stanley E. Porter and Andrew W. Pitts, 261-82. Linguistic Biblical Studies 6. Boston: Brill, 2013.
- . “Style and Pseudonymity in Pauline Scholarship: A Register Based Configuration.” In *Paul and Pseudepigraphy*, edited by Stanley E. Porter and Gregory P. Fewster, 113–52. Pauline Studies 8. Leiden; Boston: Brill, 2013.

- Polanyi, Michael. *Personal Knowledge: Towards a Post-Critical Philosophy*. New York: Harper & Row, 1964.
- . *Science, Faith, and Society*. Chicago; London: University of Chicago Press, 1964.
- Popper, Karl R. *Conjectures and Refutations: The Growth of Scientific Knowledge*. London; New York: Taylor & Francis, 2014.
- Porter, Stanley E. “The Argument of Romans 5: Can a Rhetorical Question Make a Difference?” *Journal of Biblical Literature* 110 (1991) 655–77.
- . *Idioms of the Greek New Testament*. 2nd ed. Biblical Languages: Greek 2. Sheffield: Sheffield Academic, 1994.
- . “Pauline Authorship and the Pastoral Epistles: Implications for Canon.” *Bulletin for Biblical Research* 5 (1995) 105–23.
- . *Studies in the Greek New Testament: Theory and Practice*. Studies in Biblical Greek 6. New York: Peter Lang, 1996.
- . *Verbal Aspect in the Greek of the New Testament: With Reference to Tense and Mood*. Studies in Biblical Greek 1. New York: Peter Lang, 1989.
- . “Word Order and Clause Structure in New Testament Greek: An Unexplored Area of Greek Linguistics Using Philippians as a Test Case.” *Filologia Neotestamentaria* 6 (1993) 177–205.
- Porter, Stanley E., and D. A. Carson, eds. *Biblical Greek Language and Linguistics: Open Questions in Current Research*. Journal for the Study of the New Testament Supplement Series 80. Sheffield: Sheffield Academic, 1993.
- , eds. *Discourse Analysis and Other Topics in Biblical Greek*. Journal for the Study of the New Testament Supplement Series 113. Sheffield: Sheffield Academic, 1995.
- , eds. *Linguistics and the New Testament: Critical Junctures*. Journal for the Study of the New Testament Supplement Series 168. Sheffield: Sheffield Academic, 1999.
- Porter, Stanley E., and Matthew R. Malcolm, eds. *Horizons in Hermeneutics: A Festschrift in Honor of Anthony C. Thiselton*. Grand Rapids: Eerdmans, 2013.
- Porter, Stanley E., and Matthew Brook O’Donnell. “Theoretical Issues for Corpus Linguistics and the Study of Ancient Languages.” In *Corpus Linguistics by the Lune: A Festschrift for Geoffrey Leech*, edited by A.P. Wilson et al., 119–37. Frankfurt am Main: Peter Lang, 2003.

- Porter, Stanley E., and Thomas H. Olbricht, eds. *Rhetoric and the New Testament: Essays from the 1992 Heidelberg Conference*. Journal for the Study of the New Testament Supplement Series 90. Sheffield: Sheffield Academic, 1993.
- Porter, Stanley E., and Andrew W. Pitts. "New Testament Greek Language and Linguistics in Recent Research." *Currents in Biblical Research* 6 (2008) 214–55.
- Porter, Stanley E., and Andrew W. Pitts, eds. *The Language of the New Testament: Context, History and Development*. Linguistic Biblical Studies 6. Boston: Brill, 2013.
- Porter, Stanley E., and Jeffrey T. Reed, eds. *Discourse Analysis and the New Testament: Approaches and Results*. Journal for the Study of the New Testament Supplement Series 170. Sheffield: Sheffield Academic, 1999.
- Porter, Stanley E., and Eckhard J. Schnabel, eds. *On the Writing of New Testament Commentaries: Festschrift for Grant R. Osborne on the Occasion of His 70th Birthday*. Texts and Editions for New Testament Study 8. Leiden; Boston: Brill, 2012.
- Porter, Stanley E., et al. *Fundamentals of New Testament Greek*. Grand Rapids: Eerdmans, 2010.
- Portnoy, Stephen L., and David L. Petersen. "Biblical Texts and Statistical Analysis: Zechariah and Beyond." *Journal of Biblical Literature* 103 (1984) 11–21.
- Preisach, Christine, and Lars Schmidt-Thieme. "Ensembles of Relational Classifiers." *Knowledge and Information Systems* 14 (2008) 249–72.
- Prendinger, H., and M. Ishizuka. "A Creative Abduction Approach to Scientific and Knowledge Discovery." *Knowledge-Based Systems* 18 (2005) 321–26.
- Preuss, Lucien, and Helmut Vorkauf. "The Knowledge Content of Statistical Data." *Psychometrika* 62 (1997) 133–61.
- Putniņš, Tālis, et al. "Advanced Text Authorship Detection Methods and Their Application to Biblical Texts." In *Complex Systems*, edited by Axel Bender, J1–J13. Proceedings of SPIE 6039. Bellingham, WA: Society of Photo Optical, 2006.
- Quine, W. V. "Two Dogmas of Empiricism." In *From a Logical Point of View: 9 Logico-Philosophical Essays*, 20–46. Cambridge, MA: Harvard University Press, 1980.
- Quinn, Jerome D. "Luke and the Pastoral Epistles." *Catholic Biblical Quarterly* 43 (1981) 488–90.
- . "Paraenesis and the Pastoral Epistles." *Semeia* 50 (1990) 189–210.

- Radday, Yehuda T. "Computerized Statistical Linguistics and the Problems of the Unity of the Book of Isaiah." In *Proceedings of the 5th World Congress of Jewish Studies*, edited by Pinchas Peli, 247–50. Jerusalem: World Union of Jewish Studies, 1973.
- . "Isaiah and the Computer: a Preliminary Report." *Computers & the Humanities* 5 (1970) 65–73.
- . "Two Computerized Statistical-Linguistic Tests concerning the Unity of Isaiah." *Journal of Biblical Literature* 89 (1970) 319–24.
- Radday, Yehuda T., and Moshe A. Pollatschek. "Vocabulary Richness in Post-Exilic Prophetic Books." *Zeitschrift für die alttestamentliche Wissenschaft* 92 (1980) 333–46.
- Radday, Yehuda T., and Haim S. Shore. "The Definite Article: A Type and/or Author-Specifying Discriminant in the Hebrew Bible." *Bulletin Association for Literary and Linguistic Computing* 4 (1976) 23–31.
- . *Genesis: An Authorship Study in Computer-Assisted Statistical Linguistics*. Analecta Biblica 103. Rome: Biblical Institute, 1985.
- Radday, Yehuda T., and Dieter Wickman. *The Unity of Isaiah in the Light of Statistical Linguistics*. Edited by Gérard E. Weil. Études quantitatives et automatisées 1. Hildesheim: H.A. Gerstenberg, 1973.
- . "Unity of Zechariah Examined in the Light of Statistical Linguistics." *Zeitschrift für die alttestamentliche Wissenschaft* 87 (1975) 30–55.
- Radday, Yehuda T., et al. "The Book of Judges Examined by Statistical Linguistics." *Biblica* 58 (1977) 469–99.
- , et al. "Genesis, Wellhausen and the Computer." *Zeitschrift für die alttestamentliche Wissenschaft* 94 (1982) 467–81.
- Rand, W. "Markov Models of Literary Style for Authorship Identification." *Fluctuation and Noise Letters* 2 (2002) L299–L303.
- Rao, C. Radhakrishna. *Linear Statistical Inference and Its Applications*. New York: John Wiley & Sons, 2009.
- Rasch, Dieter, et al. *Optimal Experimental Design with R*. Boca Raton: Chapman & Hall/CRC, 2011.
- Redalié, Y. *Paul après Paul: Le temps, le salut, la morale selon les épîtres à Timothée et à Tite*. Le Monde de la Bible 31. Geneva: Labor et Fides, 1994.

- Reed, Jeffrey T. "The Cohesiveness of Discourse." In *Discourse Analysis and the New Testament: Approaches and Results*, edited by Jeffrey T. Reed and Stanley E. Porter, 28–46. Journal for the Study of the New Testament Supplement Series 170. Sheffield: Sheffield Academic, 1999.
- . *A Discourse Analysis of Philippians: Method and Rhetoric in the Debate over Literary Integrity*. Journal for the Study of the New Testament Supplement Series 136. Sheffield: Sheffield Academic, 1997.
- Reichertz, Jo. "Induction, Deduction, Abduction." In *The SAGE Handbook of Qualitative Data Analysis*, edited by Uwe Flick, 123–36. London: Sage.
- Reynolds, H. T. *Analysis of Nominal Data*. 2nd ed. Quantitative Applications in the Social Sciences 7. Newbury Park: Sage, 1984.
- Rhoads, D. M. *Reading Mark: Engaging the Gospel*. Minneapolis: Fortress, 2004.
- Rickford, John R., and Faye McNair-Knox. "Addressee- and Topic-Influenced Style Shift: A Quantitative Sociolinguistic Study." In *Sociolinguistic Perspectives on Register*, edited by Douglas Biber and Edward Finegan, 235–76. New York; Oxford: Oxford University Press, 1994.
- Riedl, Hermann. *Anamnese und Apostolizität: Der Zweite Petrusbrief und das theologische Problem neutestamentlicher Pseudepigraphie*. Regensburger Studien zur Theologie 64. Frankfurt am Main; New York: Lang, 2005.
- Robbins, Vernon K. "The Dialectical Nature of Early Christian Discourse." *Scriptura* 59 (1996) 353–62.
- Robins, Robert Henry. *Diversions of Bloomsbury; Selected Writings on Linguistics*. North-Holland Linguistic Series 1. Amsterdam: North-Holland, 1970.
- . *A Short History of Linguistics*. 4th ed. Longman Linguistics Library 6. London; New York: Longman, 1997.
- Rogers, Everett Mitchell, and D. Lawrence Kincaid. *Communication Networks: Toward a New Paradigm for Research*. New York: London: Free Press; Collier Macmillan, 1981.
- Rudman, Joseph. "Non-Traditional Authorship Attribution Studies in the Historia Augusta: Some Caveats." *Literary & Linguistic Computing* 13 (1998) 151–57.
- . "On Determining a Valid Text for Non-Traditional Authorship Attribution Studies: Editing, Unediting, and De-Editing." In *ALLC/ACH Joint Conference 2003*, 18. Athens, GA, 2003.

- . “The State of Authorship Attribution Studies: Some Problems and Solutions.” *Computers and the Humanities* 31 (1997) 351–65.
- . “The State of Non-Traditional Authorship Attribution Studies-2012: Some Problems and Solutions.” *English Studies* 93 (2012) 259–74.
- Russell, Bertrand. “On Denoting.” *Mind* 14 (1905) 479–93.
- Rustagi, Mayur, et al. “Learning Age and Gender of Blogger from Stylistic Variation.” In *Pattern Recognition and Machine Intelligence*, edited by Santanu Chaudhury et al., 205–12. Lecture Notes in Computer Science 5909. Berlin; Heidelberg: Springer, 2009.
- Sammon, John W. Jr. “A Nonlinear Mapping for Data Structure Analysis.” *IEEE Transactions on Computers* C-18 (1969) 401–09.
- Sampson, Geoffrey. *Empirical Linguistics*. London: Continuum, 2002.
- . *Schools of Linguistics: Competition and Evolution*. London: Hutchinson, 1980.
- Sanders, Ed Parish. “Literary Dependence in Colossians.” *Journal of Biblical Literature* 85 (1966) 28–45.
- Sankoff, David. “Probability and Linguistic Variation.” *Synthese* 37 (1978) 217–38.
- Santini, Marina. “Riding the Rough Waves of Genre on the Web: Concepts and Research Questions.” In *Genres on the Web: Computational Models and Empirical Studies*, edited by Alexander Mehler et al., 3–30. Dordrecht; New York: Springer, 2010.
- Sapkota, Upendra, et al. “The Use of Orthogonal Similarity Relations in the Prediction of Authorship.” In *Computational Linguistics and Intelligent Text Processing*, edited by Alexander Gelbukh, 463–75. Lecture Notes in Computer Science 7817. Berlin; Heidelberg: Springer, 2013.
- Sarndal, Carl-Erik. “On Deciding Cases of Disputed Authorship.” *Applied Statistics* 16 (1967) 251–68.
- Savoy, Jacques. “Authorship Attribution Based on Specific Vocabulary.” *ACM Transactions on Information Systems* 30 (2012) 1-30.
- Schaff, Philip. *St Augustin’s the City of God and Christian Doctrine*. Edited by Philip Schaff. Translated by Rev. Marcus Dods. A Select Library of Nicene and Post-Nicene Fathers of the Christian Church II. New York: Charles Scribner’s Sons, 1907.
- Schleiermacher, Friedrich Daniel Ernst. *Ueber den sogenannten ersten Brief des Paulos an den Timotheos: Ein kritisches Sendschreiben an J. C. Gass, Consistorialassessor und Feldprediger zu Stettin*. Berlin: Realschulbuchh, 1807.

- Schmidt, J.E.C. *Historisch Kritische Einleitung ins Neue Testament*. Giessen: Tasché und Müller, 1804.
- . “Vermutungen über die beiden Briefe an die Thessalonicher.” *Bibliothek für Kritik und Exegese des Neuen Testaments und älteste Christengeschichte* 2 (1801) 380-86.
- Schoenbaum, S. *Internal Evidence and Elizabethan Dramatic Authorship: An Essay in Literary History and Method*. Evanston: Northwestern University Press, 1966.
- Schweitzer, Albert. *Geschichte der Leben-Jesu-Forschung*. Tübingen: J. C. B. Mohr, 1913.
- Sebastiani, Fabrizio. “Machine Learning in Automated Text Categorization.” *ACM Computing Surveys* 34 (2002) 1–47.
- Sebeok, Thomas Albert, ed. *Style in Language*. Cambridge, MA: MIT Press, 1960.
- Sedelow, Sally Yeates, et al. “Some Parameters for Computational Stylistics: Computer Aids to the Use of Traditional Categories in Stylistic Analysis.” In *Proceedings of the IBM Literary Data Processing Conference*, 211-29. Yorktown Heights, NY: IBM, 1964.
- Semino, Elena. “Stylistics and Linguistic Variation in Poetry.” *Journal of English Linguistics* 30 (2002) 28–50.
- Sgouros, Tom. “What Is Context for? Syntax in a Non-Abstract World.” *Journal of Logic, Language, and Information* 14 (2005) 235–51.
- Shannon, Claude Elwood, and Warren Weaver. *The Mathematical Theory of Communication*. Urbana: University of Illinois Press, 1949.
- Sherman, Lucius Adelno. “On Certain Facts and Principles in the Development of Form in Literature.” *University Studies* 1 (1892) 337-66.
- . “Some Observations upon the Sentence-Length in English Prose.” *University Studies* 1 (1888) 119–30.
- Sichel, H. S. “On a Distribution Representing Sentence-Length in Written Prose.” *Journal of the Royal Statistical Society* 137 (1974) 25–34.
- Sigelman, Lee, and William Jacoby. “The Not-So-Simple Art of Imitation: Pastiche, Literary Style, and Raymond Chandler.” *Computers and the Humanities* 30 (1996) 11–28.
- Silva, Moisés. *Biblical Words and Their Meaning: An Introduction to Lexical Semantics*. Grand Rapids: Zondervan, 1983.

- Simcox, William Henry. *The Writers of the New Testament: Their Style and Characteristics*. New York: Thomas Whittaker, 1892.
- Simpson, E. H. "Measurement of Diversity." *Nature* 163 (1949) 688.
- Simpson, Edmund Kidley. *The Pastoral Epistles: The Greek Text with Introduction and Commentary*. Wheaton: Tyndale, 1954.
- Sisson, Russell B. "Abductive Logic and Rhetorical Structure in 1 Corinthians 9." *Proceedings Eastern Great Lakes & Midwest Biblical Societies* 26 (2006) 93–100.
- Smith, Craig A. *Timothy's Task, Paul's Prospect: A New Reading of 2 Timothy*. New Testament Monographs 12. Sheffield: Sheffield Phoenix Press, 2006.
- Smith, M. W. A. "The Authorship of 'The Revenger's Tragedy'." *Notes and Queries* 38 (1991) 508–13.
- . "The Authorship of a Lover's Complaint: An Application of Statistical Stylometry to Poetry." *Computers and the Humanities* 18 (1984) 23–55.
- . "The Authorship of Acts I and II of 'Pericles': A New Approach Using First Words of Speeches." *Computers and the Humanities* 22 (1988) 23–41.
- . "Computers, Statistics and Disputed Authorship." In *Language and the Law*, edited by John Gibbons, 374–413. New York: Longman, 1994.
- . "Edmund Ironside." *Notes and Queries* 40 (1993) 202–04.
- . "An Investigation of Morton's Method to Distinguish Elizabethan Playwrights." *Computers and the Humanities* 19 (1985) 3–21.
- . "Merriam's Applications of Morton's Method." *Computers and the Humanities* 21 (1987) 59–60.
- . "The Problem of Acts I–II of Pericles." *Notes and Queries* 39 (1992) 346–55.
- . "A Procedure to Determine Authorship Using Pairs of Consecutive Words: More Evidence for Wilkins's Participation in 'Pericles'." *Computers and the Humanities* 23 (1989) 113–29.
- . "Recent Experience and New Developments of Methods for the Determination of Authorship." *Bulletin of the Association for Literary and Linguistic Computing* 11 (1983) 73–82.
- . "The Revenger's Tragedy: The Derivation and Interpretation of Statistical Results for Resolving Disputed Authorship." *Computers and the Humanities* 21 (1987) 21–68.

- . “Shakespeare, Stylometry and ‘Sir Thomas More’.” *Studies in Philology* 89 (1992) 434–45.
- . “Word-Links and Shakespearian Authorship and Chronology.” *Notes and Queries* 35 (1988) 57–59.
- . “Word-Links and the Authorship of Edmund Ironside.” *Notes and Queries* 35 (1988) 447–49.
- Smith, Peter W. H., and David A. Rickards. “The Authorship of the American Declaration of Independence.” Paper presented at the AISB2008 Convention: Symposium on Style In Text: Creative Generation and Identification of Authorship, Aberdeen, Scotland, April 1–4, 2008.
- Somers, H., and F. Tweedie. “Authorship Attribution and Pastiche.” *Computers and the Humanities* 37 (2003) 407–29.
- Song, Changwon. *Reading Romans as a Diatribe*. Studies in Biblical Literature 59. New York: Peter Lang, 2004.
- Spector, Paul E. *Research Designs*. Quantitative Applications in the Social Sciences 23. Newbury Park: Sage, 1981.
- Spencer, Aida Besançon. *Paul’s Literary Style: A Stylistic and Historical Comparison of II Corinthians 11:16-12:13, Romans 8:9-39, and Philippians 3:2-4:13*. Jackson, MI: Evangelical Theological Society, 1984.
- Spencer, F. Scott. “Preparing the Way of the Lord.” In *Reading Luke*, edited by Craig G. Bartholomew et al., 104–24. Milton Keynes, UK; Grand Rapids: Paternoster; Zondervan, 2005.
- Speyer, Wolfgang. *Die literarische Fälschung im heidnischen und christlichen Altertum: Ein Versuch ihrer Deutung*. Handbuch der Altertumswissenschaft 1.2. München: Beck, 1971.
- Spicq, Ceslas. *Saint Paul, les Épîtres Pastorales*. Paris: J. Gabalda et Cie, 1969.
- Stamatatos, Efstathios. “Authorship Attribution Based on Feature Set Subspacing Ensembles.” *International Journal on Artificial Intelligence Tools* 15 (2006) 823–38.
- . “Plagiarism Detection Using Stopword N-Grams.” *Journal of the American Society for Information Science and Technology* 62 (2011) 2512–27.
- . “A Survey of Modern Authorship Attribution Methods.” *Journal of the American Society for Information Science and Technology* 60 (2009) 538–56.

- Stamatatos, Efstathios, et al. "Automatic Text Categorization in Terms of Genre and Author." *Computational Linguistics* 26 (2000) 471–95.
- . "Computer-Based Authorship Attribution without Lexical Measures." *Computers and the Humanities* 35 (2001) 193–214.
- . "Text Genre Detection Using Common Word Frequencies." Paper presented at the Proceedings of the 18th Conference on Computational linguistics, Saarbrücken, Germany, 2000.
- Stańczyk, Urszula. "Application of DRSA-ANN Classifier in Computational Stylistics." In *Foundations of Intelligent Systems*, edited by Marzen A. Kryszkiewicz et al., 695–704. Lecture Notes in Computer Science Berlin; Heidelberg: Springer, 2011.
- . "On Performance of DRSA-ANN Classifier." In *Hybrid Artificial Intelligent Systems*, edited by Emilio Corchado et al., 172–79. Lecture Notes in Computer Science 6679. Berlin; Heidelberg: Springer 2011.
- Stein, Benno, et al. "Intrinsic Plagiarism Analysis." *Language Resources and Evaluation* 45 (2011) 63–82.
- Stephens, M. B. *Annihilation or Renewal? The Meaning and Function of New Creation in the Book of Revelation*. Wissenschaftliche Untersuchungen zum Neuen Testament 2. Tübingen: Mohr Siebeck, 2011.
- Stewart, Larry L. "Charles Brockden Brown: Quantitative Analysis and Literary Interpretation." *Literary and Linguistic Computing* 18 (2003) 129–38.
- Stolerman, Ariel, et al. "Breaking the Closed-World Assumption in Stylometric Authorship Attribution." In *Advances in Digital Forensics X*, edited by Gilbert Peterson and Sujeet Sheno, 185–205. IFIP Advances in Information and Communication Technology 433. Berlin; Heidelberg: Springer, 2014.
- Stowers, Stanley Kent. *The Diatribe and Paul's Letter to the Romans*. Society of Biblical Literature Dissertation Series 57. Chico, CA: Scholars, 1981.
- . *Letter Writing in Greco-Roman Antiquity*. Philadelphia: Westminster, 1986.
- Stratil, M., and R. J. Oakley. "A Disputed Authorship Study of Two Plays Attributed to Tirso de Molina." *Literary & Linguistic Computing* 2 (1987) 153–60.
- Streitberg, Bernd. "Exploring Interactions in High-Dimensional Tables: A Bootstrap Alternative to Log-Linear Models." *The Annals of Statistics* 27 (1999) 405–13.
- Stuart, Lauren M., et al. "On Identifying Authors with Style." In *2013 IEEE International Conference on Systems, Man, and Cybernetics*, 3048–53. IEEE International Conference

on Systems Man and Cybernetics Conference Proceedings Los Alamitos, CA: IEEE, 2013.

Stuhlmacher, Peter, ed. *The Gospel and the Gospels*. Grand Rapids: Eerdmans, 1991.

Suero Montero, Calkin, et al. "Investigating the Role of Emotion-Based Features in Author Gender Classification of Text." In *Computational Linguistics and Intelligent Text Processing*, edited by Alexander Gelbukh, 98–114. Lecture Notes in Computer Science 8404. Berlin; Heidelberg: Springer, 2014.

Svartvik, Jan, ed. *Directions in Corpus Linguistics: Proceedings of Nobel Symposium 82 Stockholm, 4-8 August 1991*. Trends in Linguistics 65. Berlin; New York: De Gruyter, 1992.

Swetnam, James. "On the Literary Genre of the 'Epistle' to the Hebrews." *Novum Testamentum* 11 (1969) 261–69.

Talbert, Charles H. *What is a Gospel? The Genre of the Canonical Gospels*. Philadelphia: Fortress, 1977.

Tallentire, D.R. "Mathematical Modelling in Stylistics: Its Extent and General Limitations." In *The Computer in Literary and Linguistics Research*, edited by R. A. Wisbey, 117–28. Cambridge: Cambridge University Press, 1971.

Tambouratzis, George. "Assessing the Effectiveness of Feature Groups in Author Recognition Tasks with the SOM Model." *IEEE Transactions on Systems Man and Cybernetics* 36 (2006) 249–59.

Tashakkori, A., and C. Teddlie. *Handbook of Mixed Methods in Social & Behavioral Research*. Thousand Oaks, CA: Sage, 2003.

Tatsuoka, Maurice M., and David V. Tiedeman. "Discriminant Analysis." *Review of Educational Research* 24 (1954) 402–20.

Taylor, Wilson L. "'Cloze Procedure': A New Tool for Measuring Readability." *Journalism Quarterly* 30 (1953) 415–33.

———. "Recent Developments in the Use of 'Cloze Procedure'." *Journalism quarterly* 33 (1956) 42-48.

Teich, Elke. *Systemic Functional Grammar in Natural Language Generation: Linguistic Description and Computational Representation*. New York; London: Cassell, 1999.

Temple, J. T. "A Multivariate Synthesis of Published Platonic Stylometric Data." *Literary and Linguistic Computing* 11 (1996) 67–75.

- Theil, Henri. *Statistical Decomposition Analysis: With Applications in the Social and Administrative Sciences*. Amsterdam: North-Holland, 1972.
- Theissen, Gerd. *The New Testament: A Literary History*. Translated by Linda M. Maloney. Minneapolis: Fortress, 2011.
- . *The New Testament: History, Literature, Religion*. London; New York: T & T Clark, 2003.
- Thiselton, Anthony C. *Interpreting God and the Postmodern Self: On Meaning, Manipulation, and Promise*. Grand Rapids: Eerdmans, 1995.
- . *New Horizons in Hermeneutics*. Grand Rapids: Zondervan, 1992.
- . *Thiselton on Hermeneutics: The Collected Works and New Essays of Anthony Thiselton*. Edited by John R. Hinnells. Grand Rapids: Eerdmans, 2006.
- Thiselton, Anthony C., and Urbanija Marko. "Thirty Years of Hermeneutics: Retrospect and Prospect." In *Interpretation of the Bible: The International Symposium in Slovenia*, edited by Jože Krašovec, 1559–74. The Library of Hebrew Bible/Old Testament Studies 289. Sheffield: Sheffield Academic, 1998.
- Thisted, Ronald, and Bradley Efron. "Did Shakespeare Write a Newly-Discovered Poem?" *Biometrika* 74 (1987) 445–55.
- Thoiron, Philippe. "Diversity Index and Entropy as Measures of Lexical Richness." *Computers and the Humanities* 20 (1986) 197–202.
- Thompson, Geoff. *Introducing Functional Grammar*. 2nd ed. London: Arnold, 2004.
- Thomson, Godfrey H., and J. Ridley Thompson. "Outlines of a Method for the Quantitative Analysis of Writing Vocabularies." *British Journal of Psychology* 8 (1915) 52–69.
- Thomson, R., and T. Murachver. "Predicting Gender from Electronic Discourse." *British Journal of Social Psychology* 40 (2001) 193–208.
- Thurstone, L. L. *The Vectors of Mind; Multiple-Factor Analysis for the Isolation of Primary Traits*. Chicago: University of Chicago Press, 1935.
- Tiao, George C. "Inference and Disputed Authorship: The Federalist. by Frederick Mosteller; David L. Wallace." *Journal of the American Statistical Association* 62 (1967) 306–09.
- Todorov, Tzvetan. "The Place of Style in the Structure of the Text." In *Literary Style: A Symposium*, edited by Seymour Chatman, 29–39. London; Oxford: Oxford University Press, 1971.

- Towner, Philip. H. "Pauline Theology or Pauline Tradition in the Pastoral Epistles: The Question of Method." *Tyndale Bulletin* 46 (1995) 287–314.
- Tracy, David. *Plurality and Ambiguity: Hermeneutics, Religion, Hope*. San Francisco: Harper & Row, 1987.
- Trobisch, David. *Paul's Letter Collection: Tracing the Origins*. Minneapolis: Fortress, 1994.
- Tschuggnall, Michael, and Günther Specht. "Using Grammar-Profiles to Intrinsically Expose Plagiarism in Text Documents." In *Natural Language Processing and Information Systems*, edited by Elisabeth Métais et al., 297–302. Lecture Notes in Computer Science 7934. Berlin; Heidelberg: Springer, 2013.
- Tsimboukakis, Nikos, and George Tambouratzis. "Neural Networks for Author Attribution." In *2007 IEEE International Conference on Fuzzy Systems, Vols 1–4*, 147–52. London: IEEE, 2007.
- . "Word-Map Systems for Content-Based Document Classification." *IEEE Transactions on Systems Man and Cybernetics* 41 (2011) 662–73.
- Tuldava, Juhan. "The Development of Statistical Stylistics (A Survey)." *Journal of Quantitative Linguistics* 11 (2004) 141–51.
- Tweedie, Fiona J. "A Statistical Investigation into the Provenance of *De Doctrina Christiana*, Attributed to John Milton." Ph.D diss., University of the West of England at Bristol, 1997.
- . "Statistical Models in Stylistics and Forensic Linguistics." In *Quantitative Linguistics / Quantitative Linguistik: An International Handbook / Ein Internationales Handbuch*, edited by Reinhard Köhler et al., 387–97. Handbücher zur Sprach und Kommunikationswissenschaft 27. Berlin; New York: de Gruyter, 2005.
- Tweedie, Fiona J., and R. Harald Baayen. "How Variable May a Constant Be? Measures of Lexical Richness in Perspective." *Computers and the Humanities* 32 (1998) 323–52.
- Tweedie, Fiona J., et al. "The Provenance of *De Doctrina Christiana*, Attributed to John Milton: a Statistical Investigation." *Literary & Linguistic Computing* 13 (1998) 77–87.
- . "Neural Network Applications in Stylometry: The Federalist Papers." *Computers and the Humanities* 30 (1996) 1–10.
- Tyson, Paul G. "Transcendence and Epistemology: Exploring Truth Via Post-Secular Christian Platonism." *Modern Theology* 24 (2008) 245–70.
- Ule, L. A. "Cluster Analysis." *Bulletin of the Association for Literary and Linguistic Computing* 2 (1974) 16–21.

- Unsworth, Anna, and A. Q. Morton. "Mrs. Gaskell Anonymous: Some Unidentified Items in 'Fraser's Magazine'." *Victorian Periodicals Review* 14 (1981) 24–31.
- Valenza, Robert J. "Are the Thisted-Efron Authorship Tests Valid?" *Computers and the Humanities* 25 (1991) 27–112.
- van Dam, Michiel, and Claudia Hauff. "Large-Scale Author Verification: Temporal and Topical Influences." In *Proceedings of the 37th international ACM SIGIR 2014 Conference on Research*, 1039–42. Gold Coast, Queensland, Australia: ACM, 2014.
- Van Gelder, Craig. "Method in Light of Scriptures and in Relation to Hermeneutics." *Journal of Religious Leadership* 3 (2004) 43–73.
- Van Halteren, Hans. "Author Verification by Linguistic Profiling." *ACM Transactions on Speech and Language Processing (TSLP)* 4 (2007) 1–17.
- van Roon, A. *The Authenticity of Ephesians*. Translated by S. Prescol-Jokel. Novum Testamentum Supplement Series 39. Leiden: Brill, 1974.
- Vanhoozer, Kevin J. *Is There a Meaning in This Text? The Bible, the Reader, and the Morality of Literary Knowledge*. Grand Rapids: Zondervan, 1998.
- Vaux, Bert. "Syllabification in Armenian, Universal Grammar, and the Lexicon." *Linguistic Inquiry* 34 (2003) 91–125.
- Viswanathan, Madhu. *Measurement Error and Research Design*. Thousand Oaks, CA: Sage, 2005.
- von Schlegel, Friedrich. *Ueber die Sprache und Weisheit der Indier: Ein Beitrag zur Begründung der Alterthumskunde*. Heidelberg: Mohr und Zimmer, 1808.
- von Weizsäcker, Karl Heinrich. *Das apostolische Zeitalter der christlichen Kirche*. Freiburg: Mohr, 1892.
- Votaw, Clyde Weber. "The Gospels and Contemporary Biographies." *The American Journal of Theology* 19 (1915) 45–73.
- Wake, William C. "Sentence-Length Distributions of Greek Authors." *Journal of the Royal Statistical Society* 120 (1957) 331–46.
- Wallace, Daniel B. *The Basics of New Testament Syntax: An Intermediate Greek Grammar*. Grand Rapids: Zondervan, 2000.
- . *Greek Grammar Beyond the Basics: An Exegetical Syntax of the New Testament: with Scripture, Subject, and Greek Word Indexes*. Grand Rapids: Zondervan, 1996.

- Wallace, Gregory K. "The JPEG Still Picture Compression Standard." *IEEE Transactions on Consumer Electronics* 38 (1992) xviii–xxxiv.
- Watson, Duane Frederick. *Invention, Arrangement, and Style: Rhetorical Criticism of Jude and 2 Peter*. SBL Dissertation Series 104. Atlanta: Scholars, 1988.
- Watson, Francis. *Text and Truth: Redefining Biblical Theology*. Grand Rapids: Eerdmans, 1997.
- Waugh, Sam, et al. "Computational Stylistics Using Artificial Neural Networks." *Literary and Linguistic Computing* 15 (2000) 187–98.
- Weingartner, Russell. "Stylistic Analysis of an Anonymous Work: The Old French Lai Guingamor." *Modern Philology: A Journal Devoted to Research in Medieval and Modern Literature* 69 (1971) 1–9.
- Wellek, René, and Austen Warren. *Theory of Literature*. 3rd ed. New York: Harcourt, Brace & World, 1962.
- Wells, Rulon "Some Neglected Opportunities in Descriptive Linguistics." *Anthropological Linguistics* 5 (1963) 38–49.
- Wenham, Gordon J. "Genesis: An Authorship Study and Current Pentateuchal Criticism." *Journal for the Study of the Old Testament* 42 (1988) 3–18.
- Whitehead, Alfred North, and Bertrand Russell. *Principia Mathematica*. London: Cambridge University Press, 1910.
- Whitelaw, Casey, et al. "Identifying Interpersonal Distance Using Systemic Features." In *Computing Attitude and Affect in Text: Theory and Applications*, edited by James G. Shanahan et al., 199–214. The Information Retrieval Series 20. Dordrecht: Springer, 2006.
- Wickmann, Dieter. "On Disputed Authorship, Statistically." *Bulletin of the Association for Literary and Linguistic Computing* 4 (1976) 32–41.
- Williams, C. B. "A Note on the Statistical Analysis of Sentence-Length as a Criterion of Literary Style." *Biometrika* 31 (1940) 356–61.
- . "Studies in the History of Probability and Statistics: IV. A Note on an Early Statistical Study of Literary Style." *Biometrika* 43 (1956) 248–56.
- Wills, Lawrence M. *The Quest of the Historical Gospel: Mark, John and the Origins of the Gospel Genre*. London: Routledge, 1997.

- Winer, George Benedict. *A Grammar of the Idioms of the Greek Language of the New Testament*. Translated by J.H. Agnew and O.G. Ebbecke. New York: Robert Carter & Brothers, 1850.
- Witten, Ian H., and Eibe Frank. *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. San Francisco: Morgan Kaufmann, 2000.
- Wittgenstein, Ludwig. *Tractatus Logico-Philosophicus*. Translated by C.K. Ogden. Mineola, NY: Dover, 1999.
- Wolter, Michael. *Die Pastoralbriefe als Paulustradition*. Göttingen: Vandenhoeck & Ruprecht, 1988.
- Young, Richard E., et al. *Rhetoric: Discovery and Change*. New York: Harcourt, Brace & World, 1970.
- Yule, G. Udny. "On Sentence-Length as a Statistical Characteristic of Style in Prose: With Application to Two Cases of Disputed Authorship." *Biometrika* 30 (1939) 363–90.
- Zhao, Ying, and Justin Zobel. "Effective and Scalable Authorship Attribution Using Function Words." In *Information Retrieval Technology, Second Asia Information Retrieval Symposium, AIRS 2005*, edited by G. G. Lee et al., 174–89. Lecture Notes in Computer Science 3689. New York: Springer, 2005.
- . "Entropy-Based Authorship Search in Large Document Collections." In *Advances in Information Retrieval*, edited by G. Amati et al., 381–92. Lecture Notes in Computer Science 4425. Berlin; Heidelberg; New York: Springer, 2007.
- Zheng, Rong, et al. "Authorship Analysis in Cybercrime Investigation." In *Intelligence and Security Informatics*, edited by Hsinchun Chen et al., 59–73. Lecture Notes in Computer Science 2665. Berlin; Heidelberg: Springer, 2003.
- Zipf, George Kingsley. *The Psycho-Biology of Language; An Introduction to Dynamic Philology*. M.I.T. Paperback Series 38. Cambridge, MA: M.I.T., 1935.
- . *Selected Studies of the Principle of Relative Frequency in Language*. Cambridge, MA: Harvard University Press, 1932.
- Zytkow, Jan M., and Robert Zembowicz. "Mining Patterns at Each Scale in Massive Data." In *Foundations of Intelligent Systems: 9th International Symposium, ISMIS'96*, edited by Z.W. Ras and M. Michalewicz, 139–48. LNAI 1079. Berlin: Springer, 1996.