MAPPING MARK: QUANTITATIVE STUDY OF CLAUSE THEMATIZATION AS A MEANS OF ILLUMINATING THE GOSPEL GENRE

by

Nathan L. Brown, BA, MA

A thesis submitted to
the Faculty of McMaster Divinity College
in partial fulfilment of the requirements
for the degree of Doctor of Philosophy (Christian Theology)

McMaster Divinity College
Hamilton, Ontario
2020

DOCTOR OF PHILOSOPHY | McMaster Divinity College
(Christian Theology) | Hamilton, ON

TITLE: Mapping Mark: Quantitative Study of Clause Thematization as a Means of Illuminating the Gospel Genre

AUTHOR: Nathan L. Brown

SUPERVISORS: Dr. Stanley E. Porter and Dr. Cynthia L. Westfall

NUMBER OF PAGES: vii + 402

McMASTER DIVINITY COLLEGE

Upon the recommendation of an oral examining committee,

this dissertation by

Nathan Brown

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

DOCTOR OF PHILOSOPHY (CHRISTIAN THEOLOGY)

Primary Supervisor: Stanley E. Porter — Digitally signed by Stanley Porter Date: 2020.05.07 15:06:28 -04'00'

Stanley E. Porter, PhD

Secondary Supervisor: Cynthia Long-Westfall — Digitally signed by Cynthia Long-Westfall Date: 2020.05.07 18:33:54 -04'00'

Cynthia Long Westfall, PhD

External Examiner: Randall Tan — Digitally signed by Randall Tan Date: 2020.05.07 13:54:54 -04'00'

Randall K. J. Tan, PhD

Vice President Academic Designate: Paul S Evans — Digitally signed by Paul S Evans DN: cn=Paul S Evans, o, ou, email=pevans@mcmaster.ca, c=CA Date: 2020.05.07 13:05:14 -04'00'

Paul S. Evans, PhD

Date: May 7, 2020

ABSTRACT

"Mapping Mark: Quantitative Study of Clause Thematization as a Means of Illuminating the Gospel Genre"

Nathan L. Brown
McMaster Divinity College
Hamilton, Ontario
Doctor of Philosophy (Christian Theology), 2020

This project exhaustively examines the first element (theme) of each clause in Mark and in samples from other roughly contemporaneous Jewish writings. The comparative documents are divided into two categories, referential and non-referential narratives. Then statistical analyses (χ^2 and t-test) are used to determine with which category of comparative documents Mark more closely aligns. The raw results of these hypothesis tests were equivocal, but their corresponding effect sizes (Cramer's *V* and Cohen's *d*, respectively) clearly demonstrate that Mark more closely resembles referential narrative, although the difference is small.

CONTENTS

LIST OF ABBREVIATIONS

ABD	Freedman, David Noel, ed. *The Anchor Bible Dictionary*. 6 vols. New York: Doubleday, 1992.
AcBib	Academia Biblica
ATANT	Abhandlungen zur Theologie des Alten und Neuen Testaments
BAGL	*Biblical and Ancient Greek Linguistics*
BDF	Blass, Friedrich, and Albert Debrunner. *A Greek Grammar of the New Testament and Other Early Christian Literature*. Translated and revised by Robert W. Funk. Chicago: University of Chicago Press, 1961.
BHL	Blackwell Handbooks in Linguistics
BibInt	Biblical Interpretation
CurBR	*Currents in Biblical Research*
CWH	Collected Works of M. A. K. Halliday
JBL	*Journal of Biblical Literature*
JSNT	*Journal for the Study of the New Testament*
JSNTSup	Journal for the Study of the New Testament Supplement Series
JSOTSup	Journal for the Study of the Old Testament Supplement Series
LBS	Linguistic Biblical Studies
LHBOTS	The Library of Hebrew Bible/Old Testament Studies
LNTS	Library of New Testament Studies
NeoT	*Neotestamentica*
NovT	*Novum Testamentum*
NovTSup	Supplements to Novum Testamenum

NTS	*New Testament Studies*
NTTS	New Testament Tools and Studies
OTP	Charlesworth, James H., ed. *The Old Testament Pseudepigrapha*. 2 vols. Garden City: Doubleday, 1983–1985.
PTMS	Pittsburgh Theological Monograph Series
QMHSS	Quantitative Methods in the Humanities and Social Sciences
SemeiaSt	Semeia Studies
SBLDS	Society of Biblical Literature Dissertation Series
SBLSBS	Society of Biblical Literature Sources for Biblical Study
SBG	Studies in Biblical Greek
SNTG	Studies in New Testament Greek
SNTSMS	Society for New Testament Studies Monograph Series
WUNT	Wissenschaftliche Untersuchungen zum Neuen Testamentum
ZNW	*Zeitschrift für die neutestamentliche Wissenschaft und die Kunde der älteren Kirche*

INTRODUCTION

Cilliers Breytenbach, as part of the conclusion of his survey of Markan studies over the first decade of the current century calls attention to the “disturbing” fact that “there is so little research on the style, syntax, and semantics” of Mark over the course of this period.[1] Earlier in his essay he mentions that for a comprehensive look at Mark’s style the main contributions are the “dated” work of M. Zerwick, C. H. Turner, and G. D. Kilpatrick.[2] Similarly, Stanley E. Porter’s essay in the same work bemoans the state of linguistic scholarship in the study of Mark: “In a discipline such as biblical studies, which has taken the art of writing long and encompassing footnotes to a new pinnacle of comprehensiveness, whereby every known or related work is cited in support or refutation, it is surprising how many comments can be made about language, without a single source—old or new—being cited.”[3] This study attempts to make a linguistically well-founded argument that Mark more closely resembles a corpus of referential narrative than a corpus of non-referential narrative and, as a result, texts like the ones constituting my corpus of referential narratives are a better analogical genre for Mark than the texts of which the corpus of non-referential narratives consists.

[1] Breytenbach, “Current Research,” 32.
[2] Breytenbach, “Current Research,” 19n34.
[3] Porter, “Matthew and Mark,” 97–98.

CHAPTER 1: SURVEY OF PREVIOUS RESEARCH

This work attempts to assess the genre of Mark's Gospel, as emblematic of the gospel genre as a whole, by quantitatively comparing the structure of its clauses to those of representative works from two candidate groups, which I term referential and non-referential narrative. This problem statement implicitly raises a number of issues, some related to the history of New Testament studies and Greek grammatical study, which I will cover here, and some methodological or procedural questions, which I will handle under those headings (Chapters 2 and 3, respectively).

Preparatory to discussing previous discussion of the gospel genre, I include a brief survey of genre studies in general and of the genre system of Hellenistic Greek. That discussion serves as background for the meat of this chapter, namely a history of previous scholarly attempts to classify the genre of the Gospels. Some of the issues raised in these surveys lead to including a study of the relationship between the Greek of the New Testament, particularly Mark, and the Greek language as a whole in general terms. Finally, the nature of my investigation necessitates a more narrowly focused discussion on the study of Greek word order and clause structure.

Background to the Question of the Gospels' Genre

This section covers two preparatory issues before I turn to discussing the history of investigations into the Gospel genre. The first issue is defining the concept of genre and how one determines the genre to which a particular work belongs. The second subsection describes some of the genres present in the Greco-Roman world. I did not choose the genres included here randomly; these are some of the genres to which previous research has compared the Gospels. They will also form the basis of the corpus to which I compare Mark in the body of this project.

Brief Overview of Genre as a Literary Concept

Discussion of genre has a long history in the study of both biblical and extrabiblical literature.[1] One can approach the concept of genre from a variety of perspectives, including both linguistics and literary criticism. Chapter 2 will discuss how various linguistic models conceive of genre as part of laying a foundation for my own project, which is essentially based in quantitative linguistics. Since I will be covering the linguistic approach elsewhere, this subsection focuses on the literary approach.

David Fishelov, a literary critic, argues that “literary genre is an elusive and multifaceted phenomenon that resists explanation by any one simple, straightforward approach.”[2] A survey of other literature on the question supports Fishelov’s assertion. Very little consensus seems to exist on the concept of genre. About the only thing about which almost everyone agrees is that the question is important.[3] Faced with finite time and space in which to communicate, authors must leave gaps in their texts and trust their readers to fill them in. Filling in these blanks incorrectly could easily distort interpretation significantly. Genre expectations are one tool authors have available to aid the reader in this process by informing the reader of “the rules of the code” governing

---

[1] Pearson and Porter, “Genres,” 131. Adams provides a survey going back to classical times (*Genre*, 26–67).

[2] Fishelov, *Metaphors*, 1.

[3] Aune (“Genre Theory,” 143) notes a few exceptions.

"how the author asks the reader to approach" the text.[4] Even authors who subvert these rules to make a point are relying on the readers' knowledge of the rules themselves.[5]

Such expectations can operate at a variety of levels, and precisely which level deserves the label "genre" is a matter of some discussion. Pearson and Porter suggest differentiating between "smaller units within complete works" and "the larger wholes of which they are constituent parts," reserving the term genre for the latter.[6]

The difficulty in defining genre has resulted in a "tendency to resort to analogy" in explaining its nature.[7] Fishelov focuses on four particularly persistent analogies: (1) genre as biological species, (2) genre as biological family, (3) genre as social institution, and (4) genre as speech act.[8] The last of these veers perilously close to linguistics and, thus, lies outside my study here, but the other three analogies each contribute something important to a proper understanding of genre as a literary concept.[9] The biological analogy reminds us that genres, like species, change over time.[10] Likewise, just as identifying a person's relatives through appearance does not require the people in question to be identical twins, the family analogy points out that members of a genre can exhibit a "*family resemblance*, according to which not all members of a genre share even

---

[4] Adams, *Genre*, 1; cf. Breytenbach, "Current Research," 32; Burridge, *Gospels*, 25–31; Porter, "Multidisciplinary," 107.

[5] Cf. Fishelov, *Metaphors*, 14; Pearson and Porter, "Genres," 133.

[6] Pearson and Porter, "Genres," 134.

[7] Fishelov, *Metaphors*, 1. Similarly, Duff ("Introduction," 1) begins his description of modern genre theory by stating: "In modern literary theory, few concepts have proved more problematic and unstable than that of genre."

[8] Fishelov, *Metaphors*, 1.

[9] Cf. Fishelov, *Metaphors*, 4–7.

[10] Fishelov, *Metaphors*, 1.

one trait."[11] The institutional analogy reminds one of genre's role in providing a "network of norms through which our experience is made culturally meaningful."[12]

Taking these analogies together, Fishelov defines genre as "a combination of prototypical, representative members and a flexible set of constitutive rules that apply to some level of literary texts, to some individual writers, usually to more than one literary period, and to more than one language and culture."[13] One of the strong points of this definition is that individual texts conform to the norms of their genre to differing degrees; it recognizes both "prototypical, representative members" and "a flexible set of constitutive rules."[14] Fishelov returns to this later in the book, referring to "hard-core" vs. peripheral members of a genre.[15]

Another way of defining genre is in terms of the structure of texts. This follows the following process: "Beginning with a corpus of texts representative of a genre within one or more social contexts, the analyst identifies common moves. A detailed analysis may count the presence of each move within the corpus aiming to identify which moves appear to be more or less obligatory and which might be considered optional or even rare. Sequences of moves are often analyzed as well, leading to the identification of common move patterns."[16]

The next question is how one decides the genre to which a particular text belongs. Tardy argues that it is not a text's "linguistic form but the rhetorical action it carries out in response to the dynamics of a social context" that determines its genre.[17] Nevertheless,

---

[11] Fishelov, *Metaphors*, 2.
[12] Fishelov, *Metaphors*, 2.
[13] Fishelov, *Metaphors*, 8.
[14] Fishelov, *Metaphors*, 8.
[15] Fishelov, *Metaphors*, 62–64.
[16] Tardy, "Genre," 56.
[17] Tardy, "Genre," 55.

she recognizes that investigations of linguistic forms can bear witness to “the rhetorical elements of genre.”[18]

This much is relatively clear. Where difficulties arise is in delineating criteria by which to assign any given work to a particular genre.[19] There are at least three perspectives on how genre functions, which Guelich terms the normative, descriptive, and interpretive, and the criteria for genre assignment depend on which perspective one adopts as primary.[20] I discuss each perspective and its merits below.

The normative function of genre sees genre as “setting the parameters within which one wrote and by which a text was critiqued.”[21] This perspective was particularly popular during classical times.[22] Taking the normative function as primary seems to entail too much rigidity in the process of genre analysis; any deviation between two texts—of which there will clearly be some—would potentially be grounds for ascribing different genres to them, since from this perspective the differences would result from differing norms. The ancients did not escape this difficulty: Burridge points out that classical authors were inconsistent in keeping their own rules.[23] If ancient writers, who I presume had cultural knowledge of the expectations associated with particular genres, were unable to apply their stated rules consistently, the idea that modern scholars working without the benefit of their extratextual knowledge would do better seems highly unlikely.[24]

---

[18] Tardy, “Genre,” 56–57.
[19] Guelich, “Gospel Genre,” 173; cf. Burridge, *Gospels*, 43.
[20] Guelich, “Gospel Genre,” 173.
[21] Guelich, “Gospel Genre,” 173.
[22] Guelich, “Gospel Genre,” 173.
[23] Burridge, *Gospels*, 54–55.
[24] Pearson and Porter (“Genres,” 136) point out that “ancient writings on generic categories should be used with great caution” because they were rules for writing, not interpreting.

The other two perspectives on genre function are more flexible and, therefore, more useful. The descriptive function focuses on arranging works into categories on the basis of similar form and/or content, and the interpretive function focuses on the effect such arrangements have on how a reader grapples with a text.[25]

In some ways the interpretive perspective falls prey to the same problem from the other direction.[26] That is to say, neither the normative nor interpretive perspectives focus on the data to which modern scholars have access, namely the ancient text themselves, but the descriptive perspective does, so it should be the basis for genre analysis by modern scholars. This subject will come up again in Chapter 2 as I discuss my methodology, but for now I turn to describing relevant portions of the genre system in Hellenistic Greek.[27]

Genre in the First Century

Having accepted the descriptive position, in which an interpreter assigns two works to the same genre on the basis of similar form and/or content, the question naturally arises: from whence do these categories arise? I will lay out my perspective on this question more fully in Chapter 2, where I lay out the linguistic model upon which my work is based. Meanwhile, for the sake of argument, I will simply adopt the perspective that genre is a cultural system of which individual humans making up said culture gain knowledge through exposure to instances of the various categories constituting the system. Graham Stanton illustrates this point with an intriguing thought experiment: if the second-century Roman library at Ephesus had received a codex of the Gospels, where "should the

[25] Guelich, "Gospel Genre," 173.
[26] Cf. Fishelov, *Metaphors*, 10–13.
[27] Guelich, "Gospel Genre," 173; cf. Adams, *Genre*, 3.

Gospels be put? Next to the histories, or the novels, or the biographies, or the religious treatises?"[28] The importance of this question rests on the fact that "a provisional decision about its literary genre must be made before we can begin to appreciate it."[29]

One difficulty that would have faced the librarian of Stanton's illustration more acutely than a modern scholar is the relationship between biography and history. In ancient times, at least some writers made a distinction between the two (e.g. Polybius and Plutarch), whereas in modern times biography is often seen as a subgenre of history.[30] The modern perspective—if indeed there is a real distinction to be made—offers several advantages. First, few ancient biographies survive, particularly from before the time of the New Testament, whereas a number of ancient histories, going back to the fifth century BC are available.[31] Second, the two genres seem to have sprung up at roughly the same time, and histories used biographies as sources, indicating that—even if, as Momigliano argues, "biography was never considered history in the ancient world"—there is some relation between the two.[32]

In contrast to history and biography, which the ancients referred to as such, designating a first-century work as "novel" or "romance" are external descriptors.[33] According to the classicist Niklas Holzberg, the ancient novel is: "an entirely fictitious story narrated in prose and ruled in its course by erotic motifs and a series of adventures which mostly take place during a journey and which can be differentiated into a number of specific fixed patterns. The protagonists or protagonist live(s) in a realistically

[28] Stanton, *Gospel Truth*, 136.
[29] Stanton, *Gospel Truth*, 136–37.
[30] Momigliano, *Development*, 1; cf. Aune, "Greco-Roman," 107–26; Dormeyer, *NT among Writings*, 220.
[31] Momigliano, *Development*, 8–10.
[32] Momigliano, *Development*, 12.
[33] Hock, "Novel," 127.

portrayed world . . . the actual characters, however, are given idealistic or comic-realistic features."[34] Hock differentiates the novel from myth and fable and history alike: it is "not so imaginative as to be fabulous, as is the case with the myth or fable; and it is realistic but only in the sense that its actions could have happened, not that its actions actually happened, as is the case with history."[35]

Genre did not—and, for that matter, does not—consist of airtight categories. Tomas Hägg speaks of "cross-fertilization" between the novel and biography, for example.[36]

The History of the Gospel Genre

The canonical Gospels, particularly the Gospel of Mark, provide an intriguing case study showing how perceptions of a work's genre affect its interpretation. Despite a history of interpretation spanning two millennia, scholars have reached no consensus on the place of the Gospel genre within the genre system of the first-century Greco-Roman world.[37] In fact, Pearson and Porter claim that the relation of the Gospels to the surrounding milieu has "been the most hotly contested" such question in New Testament studies.[38] This situation has led some recent scholars to despair of an answer to this question and to suggest that perhaps the difficulty stems from the question being fundamentally misconceived: what if the canonical Gospels do not share a genre or—perhaps better stated—only came to share a genre as a result of being canonized as a group?[39] However,

---

[34] Holzberg, *Ancient Novel*, 26–27.
[35] Hock, "Greek Novel," 133.
[36] Hägg, *Art of Biography*, 2.
[37] Diehl, "Gospel," 172.
[38] Pearson and Porter, "Genres," 137.
[39] Cf., e.g., Cook, *Structure*, 302–3; Diehl, "Gospel," 183–84. Tomas Hägg (*Art of Ancient Biography*, 150) seems to belong here as well, as focused on the idea that a biography covers the entirety of its subject's life, he concludes: "Narratives of a couple of years in a person's life are not necessarily to be

this historical survey will stick to considering the Gospels' genre together for two reasons. First, this particular *a priori* has much to commend it, as the discussion below will show. Second, the purpose of this survey is to survey past work, which has, almost without exception, assumed that the Gospels share a genre.[40]

The history of research into the Gospel genre consists of three phases: (1) an initial stage presupposing both the viability of analogies from the contemporary milieu and which potential analogy was correct, (2) a middle stage lasting for much of the twentieth century that questioned the viability of analogies under the influence of form criticism, and (3) a final stage, beginning around 1970, in which the crumbling of the form-critical consensus reopened the question of analogies, albeit in a more methodologically aware manner. At the moment, the preponderance of scholarly opinion seems to favor the analogical approach, although some still maintain the *sui generis* perspective associated with form-criticism and even proponents of the analogical approach present a less united front than their forebearers during the initial stage.[41] I now turn to covering each of these stages in detail.

Stage 1: The Gospels as Self-Evident Biographies

For over a millennium and a half, most interpreters of the Gospels did not question their status as biographies of Jesus Christ.[42] As early as the middle of the second century, Justin Martyr compares the Gospels to Xenophon's *Memorabilia*.[43] This was the case

classified as biographies, and there is evidently more that combines than distinguishes the four gospels; *ergo*, they are all 'gospels', but not biographies." By way of rejoinder, I might point out that, even if "narratives of a couple of years" are not "necessarily biography," they are not necessarily disqualified from being biographies, either.

[40] Petersen ("Gospel Genre," 137–58) is an example of an exception.

[41] Cf. Burridge, *Gospels*, 3.

[42] Dormeyer, *NT among Writings*, 223.

[43] Justin Martyr, *Dialogue*, 99–107, cited by Guelich, "Gospel Genre," 178n35.

despite Jesus being an unusual subject for a biography. Even into the nineteenth century, scholars involved in the original Quest for the historical Jesus treated the Gospels as ancient biographies from which they could pull the material for a biography in the modern sense.[44]

Clyde Votaw, whose work overlaps chronologically with the beginnings of the second stage, represents the height of early comparisons between Greco-Roman biographies and the Gospels. Dawson points out that comparative methodologies were the prevailing paradigm of the day.[45] However, subsequent scholarship has exposed four shortcomings of Votaw's work that future work needs to take into account.[46] First, Burridge criticizes Votaw for insufficiently attending to the difficulties in defining genre: he discusses only content and purpose, neglecting the means by which the author communicates that content and achieves that purpose.[47] Second, Guelich notes how scholars have often missed how broadly Votaw defined biography: "broadly defining biography as a work about persons and their message" allowed Votaw to cite works that a stricter definition would have excluded.[48] Third, Votaw cast too wide of a net chronologically. His suggested parallels for the Gospels range in date from the fourth century BCE (Plato/Xenophon) to the third century CE (Philostratus' *Life of Appollonius*). A timespan of seven centuries between the earliest and latest documents raises serious questions about how comparable they are to each other (let alone the Gospels). Finally, Votaw seems to have viewed ancient biographies through the lens of

---

[44] Becker, "Mark in Context," 124; cf. Walton, "Burridge's Impact," 82.
[45] Dawson, "Unmasking Flawed Consensus," 35.
[46] Cf. Schmidt, *Place*, 3–4.
[47] Burridge, *Gospels*, 5.
[48] Guelich, "Gospel Genre," 179.

modern biography.[49] Now I turn to surveying the second major period of research into the Gospel genre.

Stage 2: The Heyday of the Derivational Approach

Whereas the first stage took for granted that valid analogies exist between the Gospels and literature from their contemporary milieu, the second stage adopted the position that the Gospels evolved out of early Christian worship. This stage roughly began around the turn of the twentieth century and continued until about 1970, although there were important precursors. Three of the major figures during this period were the giants of German form-criticism: Karl-Ludwig Schmidt, Martin Dibelius, and Rudolf Bultmann.[50]

The consensus that the Gospels should be studied independently from Greco-Roman literature arose largely because of a negative value judgment of the literariness of the Gospels on the part of scholars like Eduard Norden, Paul Wendland, and—perhaps most influentially—Franz Overbeck.[51] Vielhauer points out that the German name of form-criticism, "Formgeschichte," which Dibelius introduced, is a "lightly modified version" of a term Overbeck and Norden used.[52] The work of these scholars is summarized here. I will discuss the question of how the New Testament in general, and Mark in particular, relates to Greek more broadly in the next major section.

Edward Norden's two-volume work *Die Antike Kunstprosa: Vom VI Jahrhundert v. Chr. bis in die Zeit der Renaissance* lives up to its name, covering artistic prose in Greek and Latin from the sixth century BC to the Renaissance.[53] This is an older work, as

[49] Cf. Burridge, *Gospels*, 60.
[50] Cf. Vielhauer, *Geschichte*, 281.
[51] Gundry, "Symbiosis," 18–19.
[52] Vielhauer, *Geschichte*, 281.
[53] Norden, *Antike Kunstprosa*, passim.

one of the footnotes demonstrates by mentioning that Norden had written the section on the Gospels and Acts before he had a chance to read Blass's *Grammatik*, which—after many editions and a translation into English—has been the standard reference grammar in New Testament studies for nearly a half century.[54] As might be expected, given its age, Norden's work displays no awareness of the papyri finds in Egypt and the consequent revaluation of the nature of post-classical Greek, so he measures the New Testament documents against the standard of classical Greek, concluding that the Gospels specifically "stand completely apart from the moderately artistic prose."[55] In another place he adds that their prose "had to be hurtful on the formal sense of the ancient reader."[56] Thus, Norden concludes that the Gospels are not literature, as is bound to be the case if one insists on applying the norms of classical Greek as a prescriptive standard.

Wendland held that "Mark minted no theological stance," and that providing the bare minimum of a storyline to string together the units of free-floating tradition that he had received constituted the extent of Mark's creative role as an author.[57] Such a perspective hardly makes Mark out to be a literary genius. Overbeck classified all Christian literature before the late second century as *Urliteratur*, i.e. an ancestor to literature, rather than literature itself.[58] Strictly speaking, these judgments on the literariness of the Gospels are independent of the question of whether or not the authors modeled the Gospels on contemporary literature. After all, even if one accepts the

[54] Norden, *Antike Kunstprosa*, 2:480 n 1; cf. Rydbeck, "What Happened," 424.
[55] Norden, *Antike Kunstprosa*, 2:480.
[56] Norden, *Antike Kunstprosa*, 2:458.
[57] Gundry, "Symbiosis," 19n5, citing Wendland, *Literaturformen*, 201. Gundry quoted the German; the English translation is my own.
[58] Cf. Cancik, "Die Gattung," 88–90.

premise that only works that aspire to be literature themselves emulate literature, the evangelists could have attempted to produce literature and failed.[59]

Nonetheless, among the form-critics who dominated New Testament studies in the early part of the twentieth century the consensus developed that the Gospels were not literature, had not intended to be literature, and were, therefore, to be studied independently of works of literature. Here I will discuss the work of three major form critics, Martin Dibelius, Karl-Ludwig Schmidt, and—the most influential—Rudolf Bultmann.[60]

Dibelius begins *From Tradition to Gospel*, his major work on form criticism of the Gospels, with this statement: "There is a theory that the history of literature is the history of its various forms. This may be true of literature properly so-called, but it cannot be applied indiscriminately to every kind of writing. It has, however, special significance when applied to materials where the author's personality is of little importance."[61] Given his attachment to studying the Gospels through their component forms, it then becomes incumbent upon Dibelius to substantiate that the Gospels fall into the category of "materials where the author's personality is of little importance" or that—as he puts it later in the paragraph—"no creative mind . . . impressed it with his own personality."[62] This demotion to mere assemblers of tradition, on which Burridge comments, is thus a methodologically necessary step for Dibelius.[63] Dibelius attempts to

[59] This appears to have been the earlier position of Graham Stanton (cf., e.g., *Jesus*, 118), but he seems to have changed his mind later in his career (e.g. *Gospel Truth*, 139). Breytenbach (*Nachfolge*, 68) may belong here as well. Interestingly, this was how Schmidt (*Place*, 5) characterized Diogenes Laertius, but he seems completely unwilling to entertain a similar classification for the Gospels.

[60] Cf. Dvorak, "Dibelius and Bultmann," 257.

[61] Dibelius, *From Tradition*, 1.

[62] Dibelius, *From Tradition*, 1.

[63] Burridge, *Gospels*, 7–8.

take this step by asserting a rigid demarcation between a “lower stratum” of published writing “which accords no place to the artistic devices and tendencies of literary and polished writing” and self-conscious literature.[64] A further consequence of this perspective is that “no connected narrative” such as a biography would have existed prior to the Gospels; according to him, the passion narrative is an “exception.”[65]

This appears to be a development of Overbeck’s work, though the English translation does not explicitly cite him at this point.[66] Dibelius recognizes that Papias, our earliest external witness to the formation of the gospels, already accords a creative role to the writers of the Gospels, but he simply dismisses the value of Papias’ testimony.[67] Having demolished—to his own satisfaction, at least—the idea that the Gospels were self-conscious literature, and appealing to Deissmann’s comparison of Paul’s letters to the documentary papyri, Dibelius sees a clear road ahead for evaluating “primitive Christian writing according to its own laws.”[68] Since he is unwilling to grant the evangelists the status of self-conscious authors and, therefore, cannot find laws in that which is unique to the writings in question, Dibelius must derive them from what is common to the community as a whole, which is why he refers to “the style which it is our part to observe as ‘a sociological result.’”[69]

Dibelius assumes that Christians expected an imminent Parousia and concludes from this that they had “neither the capacity nor the inclination” for literary pursuits.[70] At

[64] Dibelius, *From Tradition*, 1–2; cf. Burridge, *Gospels*, 11.
[65] Dibelius, *From Tradition*, 178.
[66] I specify the English translation because Woolf, Dibelius’s translator states (*From Tradition*, xi–xii) that Dibelius allowed him leeway in including or excluding footnotes according to his judgment as to whether they would be helpful to an English reader.
[67] Dibelius, *From Tradition*, 3–4.
[68] Dibelius, *From Tradition*, 6.
[69] Dibelius, *From Tradition*, 7.
[70] Dibelius, *From Tradition*, 9.

the root of Dibelius' approach are two factors: (1) explaining why the early Christians preserved the particular sayings and narratives about Jesus found in the Gospels, rather than others they could have preserved, and (2) discovering the "*the law* which governed" the transmission of tradition about Jesus in order for the tradition to be maintained intact.[71] His search for both of these depends on his prior assertion that the evangelists are not literary authors: "If there is no such law, then the writing of the Gospels implies not an organic development of the process by means of collecting, trimming, and binding together, but the beginning of a new and purely literary development" and without "such motive, then it is quite impossible to understand how men who made no pretensions to literature should create a tradition which constituted the first steps" towards literature.[72]

Although Dibelius is unwilling to allow for any external parallels for the entire shape of a gospel, he draws his understanding of the individual forms of which he claims the gospels are constituted almost entirely from some Jewish but mostly Greco-Roman sources.[73] Besides rabbinic anecdotes, he notices three sorts of Greco-Roman analogies: "traditions concerning the sayings of famous men" that could enlighten and encourage, "tales," and "miracle records, epiphanies and aretologies preserved in connection with some religious cultus."[74] Dibelius compares the collections of sayings, known as "Chriae" with the Gospels in the following way: "There is a similarity of *origin*, a wide difference of *content*, which influences the diction, and a certain but essential difference of *construction*."[75]

[71] Dibelius, *From Tradition*, 11.
[72] Dibelius, *From Tradition*, 11.
[73] Dibelius, *From Tradition*, 133–72.
[74] Dibelius, *From Tradition*, 151.
[75] Dibelius, *From Tradition*, 156.

Dibelius developed Overbeck's concept of *Urliteratur* into a rigid schema that classified the Gospels with popular folk writings developed from oral tradition (*Kleinliteratur*) instead of consciously-developed literature (*Hochliteratur*).[76] This rigid dichotomy causes more problems than it solves. As Reiser shows, even if one accepts that Mark falls into the category of folk literature derived from oral tradition, one cannot escape the evident "superiority of the evangelist in the processing of tradition, in narrative art, composition, and representational craft" of Mark, the "quality" of whose work Reiser credits as being a sizable factor in the success of Christianity.[77] In keeping with the attitude of his times, Dibelius' work very clearly displays an evolutionary tendency in which "the faith of Christendom moved from its fundamental strangeness in the world and its self-limitations to the religious interests of the Church, to an accommodation to the world and to harmony with its relationships."[78]

Schmidt, who was responsible for the famous description of the Gospels as "pearls on a string," devoted far more attention to the shape of individual pearls (i.e. units of tradition) than he did on the shape of the whole necklace (i.e. the Gospel itself).[79] He demoted the evangelists to mere assemblers of tradition.[80] Schmidt was willing to allow that the Gospels "do represent biography of some sort," but his attachment to classifying the Gospels as *Kleinliteratur,* rather than *Hochliteratur*, e.g. Greco-Roman biographies, forced him "to clarify the essence of ancient biography," which turns out to mean

---

[76] Burridge, *Gospels*, 11. Regarding this dichotomy, Aune helpfully states that it "was an artificial distinction that owed more to romantic notions of primitivity than to insights into comparative literature" ("Genre Theory," 150n23).

[77] Reiser, "Der Alexanderroman," 161.

[78] Dibelius, *From Tradition*, 287.

[79] Cf. Dawson, "Unmasking Flawed Consensus," 36.

[80] Burridge, *Gospels*, 7–8; cf. Dibelius, *From Tradition*, 177.

repeatedly asserting the distinction between *Kleinliteratur* and *Hochliteratur*.[81] Schmidt's rules for reconstructing the original shape of the pearls that constitute the necklace of the Gospels, which are a means for "sharpening the eye for that which is unique to primitive Christianity," depend on the presupposition that primitive Christian works had no pretensions to secular respectability; for Schmidt, this becomes "a theological matter," in which a "philosophical question about content and form is transferred into a theological question about God and the world, about Christianity and culture."[82]

Schmidt's unwavering devotion to the distinction between *Kleinliteratur* and *Hochliteratur* manifests itself repeatedly in his response to Votaw and others who stuck with the biographical perspective. He dismisses Votaw's parallels with Xenophon, Arrian, and Philostratus because these biographers "reveal specific authorial personalities," implying that the evangelists do not.[83] Similarly, although he agrees that Heinrici has found a valid point of formal comparison between the compositional process used by Diogenes Laertius and others in composing lives of philosophers and the process supposedly used for the Gospels, he questions the inference Heinrici draws from it: "The same standard of judgment cannot possibly be applied to both the gospels and Diogenes Laertius, since he tries to pass himself off as an author."[84] Furthermore, the Alexandrian biographers "sought to be scholarly," so "they must be judged differently from the evangelists."[85] He rejects Zahn's appeal to the testimony of Justin Martyr by stating that Justin "is determined to elevate the cultural level of Christianity, and to that end he

[81] Schmidt, *Place*, 3.
[82] Schmidt, *Place*, 86.
[83] Schmidt, *Place*, 4.
[84] Schmidt, *Place*, 5.
[85] Schmidt, *Place*, 5.

employs the designation 'memoirs' to locate the gospels in the high literature."[86] Schmidt is convinced that the Gospels neither are nor intended to be literature, so he is forced to conclude that Justin is ascribing to them a status they did not previously have. Schmidt concludes the discussion by simply asserting that the appeal Dibelius makes to the delay of the Parousia as forcing Papias, Justin, and their contemporaries is "the only conclusion that can possibly be reached, however, if the gospels are linked with Xenophon's *Memorabilia*" because, once again, "[l]ow literature is simply not the same as high literature."[87] It is, indeed, the only conclusion to draw if one adopts Overbeck's thesis as a premise, but perhaps one could conclude instead that it is Overbeck who was wrong.

This latter discussion sets the stage on which it becomes clear that Schmidt is driving a wedge between the Gospels and the earliest eyewitnesses.[88] According to Schmidt, Zahn focused on Justin characterizing the Gospels as memoirs because he saw in this testimony "the surest confirmation of the reliability of the Gospels, since it means that they must have actually been 'memoirs' of the apostles, just like Xenophon's 'memoirs' of Socrates."[89] Schmidt does not treat Zahn's argument fairly: he takes Zahn to task for failing to recognize that modern scholars "have to distinguish between tradition and composition" when the last sentence he quotes from Zahn on the previous page explicitly acknowledges this and argues that the same was true for Xenophon.[90] The point at issue is not whether a layer of oral tradition lies behind the written Gospels but what sort of process transmitted that oral tradition.

[86] Schmidt, *Place*, 8.
[87] Schmidt, *Place*, 10.
[88] Schmidt, *Place*, 8–10.
[89] Schmidt, *Place*, 8, citing Zahn, *Geschichte*, 471.
[90] Schmidt, *Place*, 9.

Rudolf Bultmann, easily one of the most influential New Testament scholars of his generation, lent that influence to establishing the evolutionary hypothesis as the default position within critical orthodoxy. He concludes his *History of the Synoptic Tradition* with these words: “So it is hardly possible to speak of the Gospels as a literary genus; the Gospel belongs to the history of dogma and worship.”[91] Bultmann gives two reasons for this conclusion: 1) the Gospel has no history as a literary form, and 2) the Gospels are an intra-Christian phenomenon.[92] As Cancik points out, however, this presupposes that the Gospels do not form a sub-category of a recognizable Greco-Roman genre and that other contemporary religious texts do not throw light on the Gospels.[93] These presuppositions more or less equate to begging the question.

Several scholars have pointed out that Bultmann’s arguments for rejecting biographies as an analogical genre rest on a misunderstanding of ancient biographical writing.[94] And, for that matter, Dibelius commits the same mistake when he states that the narrator of Mark is not concerned with “biographical material” and then gives a list of things that are indeed features of modern biography but not necessarily ancient ones.[95] Schmidt does likewise.[96]

Güttgemanns seems to imply that accepting Overbeck’s judgment regarding the literariness of the Gospels allowed an environment in which flights of tradition history could flourish.[97] Güttgemanns also critiques form criticism and its successors because

---

[91] Bultmann, *History*, 374; cf. Walton, “Burridge’s Impact,” 82–83.
[92] Bultmann, *History*, 374.
[93] Cancik, “Die Gattung,” 88nn14–15; cf. Güttgemanns, *Candid Questions*, 18–20.
[94] Cancik, “*Bios*,” 129; Stanton, *Gospel Truth*, 137–38. However, in fairness, it should be stated that the classicist Momigliano is closer to Bultmann than Cancik and Stanton (*Development*, 11).
[95] Dibelius, *From Tradition*, 49.
[96] Schmidt, *Place*, 11.
[97] Güttgemanns, *Candid Questions*, 1–2.

they preferred ignoring problems with their reconstructions to fixing them.[98] One example of this, which I noticed in my own reading of Dibelius, is his argument for development in the tradition with regard to the sons of Zebedee requesting pre-eminence in the Kingdom and Jesus' rebuke thereof (Mk 10:37–40): Dibelius argues for an earlier tradition that did not specify the disciples by name on the grounds that the final form of the text "shows a real interest in the significance of the sons of Zebedee, although this interest is surely foreign to the Paradigm."[99] In other words, Dibelius asserts that this pericope is a Paradigm and uses his definition of this form as grounds for altering the passage, rather than altering his description of a paradigm or the pericope's assignment to that form.

In terms of assessing the assumptions that led to Dibelius and Bultmann's conclusions, James D. Dvorak points to the work of William Wrede as "a key factor that fertilized the soil from which *Formgeschichte* sprouted."[100] What Breytenbach says with regard to Wrede is true of many others: "Wrede presupposes that, if one wants to understand literature, understanding the text as *history* must be renounced . . . [and] it is therefore not allowed to look for historical connections or links in the Gospel of Mark. Whoever intends to do this does not understand Mark's narrative manner."[101]

---

[98] Güttgemanns, *Candid Questions*, 2–3.
[99] Dibelius, *From Tradition*, 51.
[100] Dvorak, "Dibelius and Bultmann," 258.
[101] Breytenbach, *Nachfolge*, 17–18. While this consensus was widespread, it was not universal. For example, Millar Burrows ("Transitions," 119–20) noted in 1929, "The exponents of *Formgeschichte* as applied to the Synoptic Gospels would have us believe that the whole framework in which Mark presents the life of Jesus, including practically all indications of time, place, and circumstance, are his own creation, only the disconnected elements having been derived from tradition . . . but there are grave objections to such a confident assumption."

In summary, this second stage of research produced a scholarly climate in which discussions about the Gospel genre were largely "precluded in advance."[102] Despite what one scholar calls "strong words and unclear concepts" substituting for evidence, the influence of the form critics was sufficient to cement the evolutionary approach in place until the rise of redaction criticism after World War II.[103] Predominant among these vociferously argued but poorly defined concepts was the presupposition that the Gospels differed in kind, rather than degree, from contemporary literature.[104] After all, as Holzberg recognizes with regard to the ancient novel, even the attempt to create a new genre builds upon genres already in place, such as—in the case with which Holzberg is concerned—the ancient novel growing out of historiography.[105] Clearly, there was a theological axe being ground in this stage; Dibelius explicitly states: "The first understanding afforded by the standpoint of Formgeschichte is that there never was a 'purely' historical witness to Jesus. Whatever was told of Jesus' words and deeds was always a testimony of faith as formulated for preaching and exhortation in order to convert unbelievers and confirm the faithful. What founded Christianity was not knowledge about a historical process, but the confidence that the content of the story was salvation."[106]

Subsequent research has recognized that the evangelists (and, for that matter, the other writers of the New Testament) display various levels of proficiency in their use of Greek, so the rigid dichotomy between *Kleinliteratur* and *Hochliteratur* would seem to

---

[102] Burridge, *Gospels*, 11; cf. Dawson, "Unmasking Flawed Consensus," 36.
[103] Cancik, "Die Gattung," 88.
[104] Burridge ("About People," 120) characterizes the idea of the Gospels being *sui generis* as "nonsense."
[105] Holzberg, *Ancient Novel*, 35.
[106] Dibelius, *From Tradition*, 295.

be insufficiently nuanced; we shall return to this question below.[107] With the support undergirding form-criticism's adoption of the evolutionary approach removed, the way was clear for returning to the search for appropriate analogical genres within the contemporary literary milieu.

Stage 3: The Tangled Resurgence of the Analogical Approach

The rise of redaction criticism (and literary criticism later) made holding onto the form-critical portrait of mindless recorders of the oral tradition impossible: if the evangelists shaped the traditions available to them in order to emphasize distinctive theologies, as redaction criticism posits, then they clearly had the ability to shape their work according to an overall plan, and so scholarship had to face the possibility that said plan came about by way of an analogy to a pre-existing literary type.[108] Put another way, once it became clear that form-critics like Schmidt had "unjustly called the evangelists collectors and compilers, it was rightly asked" if their portrait of the place of the Gospels "in literature-history based on this classification need not be corrected."[109] Over the course of the 1970's and into the 1980's the varying results emerging from the results of scholars using the same tools made the subjectivity of redaction-criticism, to say nothing of the other historical-critical tools, clear, leading to a fracturing of the tradition-historical consensus.[110]

Once the scholarly climate thusly became more favorable, a number of different potential analogies emerged—a far greater variety, in fact, than during the previous reign

[107] Cf., e.g., Burridge, "About People," 138–40.
[108] Diehl, "Gospel," 174–75; Walton, "Burridge's Impact," 85.
[109] Breytenbach, *Nachfolge*, 68.
[110] Cf. Breytenbach, *Nachfolge*, 12–13; Güttgemanns, *Candid Questions*, 9–10.

of the analogical approach. Some of these primarily apply to other Gospels.[111] Others explain only parts of Mark, but not its overall shape.[112] I concentrate here on history and biography—similar to the first stage of research—and the ancient novel, since those are the genres to which I will compare Mark.[113]

Perhaps the most prevalent stream of research in this third period is the revival of Votaw's hypothesis that the Gospels emulate Greco-Roman biographies or, as Aune suggests, a parody thereof.[114] One recent survey of research into the genre of the Gospels argues that "scholarship largely adopted" this position.[115] This potential analogy has gained sufficient support that one scholar interested in the question of the audience(s) the evangelists may have had in mind in composing the Gospels took their biographical nature as a given and moved on to a subsequent question: "*Why βίος?*"[116] This move seems somewhat premature, but that he can make this move at all indicates the influence of this hypothesis, though even Smith has to admit that detractors remain.[117]

Taking note of the criticism scholars have levied against Votaw's work, a number of these scholars have attempted to define biography more narrowly. Lührmann, for example, tried to trace a sub-genre of biography focusing on its subject's life and death, but Guelich appropriately responds that this definition leaves too many parts of Mark's text unexplained so that instead of "a genre from which to interpret the parts we seem to have a constituent part that has been inappropriately defined as the genre of the

[111] E.g. Blomberg's suggestion that Matthew is a midrash (cf. Diehl, "Gospel," 177) or Alexander's suggestion that Luke's preface corresponds to those of scientific treatises (cf. Alexander, "Luke's Preface," 48–74).
[112] E.g. comparisons to the OT and apocalypses (Guelich, "Gospel Genre," 176–78).
[113] Cf. Stanton, *Gospel Truth*, 138.
[114] Burridge, *Gospels*, 22–23, 78–101; cf. Aune, "Genre Theory," 167–69.
[115] Walton, "Burridge's Impact," 82.
[116] Smith, *Why βίος*, passim.
[117] Smith, *Why βίος*, 5–10.

whole."[118] Additionally, the parallels are somewhat tenuous. Shuler attempted to identify the Gospels with "laudatory biography." Burridge does not find Shuler's parallels convincing.[119] More promisingly, Dormeyer argues that one should see the Gospels as bringing Greco-Roman biography into "relation with idyllic Old Testament biographical narratives."[120] These arguments failed to convince the majority of scholars.

A final suggested parallel is the ancient novel.[121] Tolbert recognizes that all of the ancient Greco-Roman novels that exist in complete form are erotic in nature, but she delineates four sorts of similarities between the Gospel of Mark and these novels: 1) mixing the form of history with the punch of drama, 2) "synthesizing" of earlier genres, 3) an episodic plot "with minimal introduction, central turning point, and final recognition scene," and 4) a "fairly crude, repetitious, and conventionalized narrative."[122]

Vines prefers to compare Mark to ancient Jewish novels, using Mikhail Bakhtin's idea of "chronotope."[123] "Chronotope" refers to the aesthetic values behind the author's aesthetic use of time and space.[124] Vines argues that Jewish novels like Judith possess a chronotope characterized by a "time of crisis and of a hostile and ironic space."[125] He then argues that the Gospel of Mark is similarly characterized, though he admits several differences (such as with regard to Jewish purity).[126] Focusing on a nebulous factor like chronotope at the explicit expense of formal similarity seems to be a weakness of the

---

[118] Guelich, "Gospel Genre," 177.

[119] Burridge, *Gospels*, 83–86; cf. Pearson and Porter, "Genre," 141.

[120] Dormeyer, *NT among the Writings*, 220–227; Dormeyer, "Evangeliumsbiographie," 35, cited by Becker, "Mark in Context," 126. Becker quoted the German; the English translation is my own.

[121] E.g. Reiser, "Der Alexanderroman," 131–63; Tolbert, *Sowing*, 62–78; Vines, *Problem*, passim; cf. Stanton, *Gospel Truth*, 139.

[122] Tolbert, *Sowing*, 65; cf. Reiser, "Der Alexanderroman," 158.

[123] E.g. Vines, *Problem*, 153; cf. Duff, "Concepts," x.

[124] Vines, *Problem*, 60–61.

[125] Vines, *Problem*, 144–52. The quotation is from page 156.

[126] Vines, *Problem*, 153–59

approach Vines takes; additionally this emphasis makes assessing the contribution his work makes to the history of research difficult, since he is essentially answering a different question.[127] Aune also raises the question of how well Vines' approach fits the majority of modern genre theory.[128] Nevertheless, his work does bring an important group of documents more clearly to view; they clearly deserve representation within a corpus of potential analogies.

To summarize, recent research into the Gospel genre has largely returned to the analogical approach that dominated the first stage of research. Despite sharing a common conviction that the Gospels belong within the contemporary literary milieu, the first and third stages differ greatly in other respects, particularly variety and methodological awareness.[129] Whereas researchers in the first stage worked with a pre-understanding that the Gospels were biographies, the third stage has produced a wide variety of suggested analogies, and both those reviving traditional hypotheses and those introducing new ones recognize the need to defend—rather than assume—their validity. Nonetheless, a complete morass has resulted in which scholars not only disagree about which potential analogy is most viable but even on the criteria by which to make that judgment.

Previous Quantitative Attempts to Study the Gospel Genre

The above survey of the resurgence of literary-historical analogies for the gospel genre purposefully omitted the work of two scholars, one of whom has been particularly influential, in order to single them out for specific attention here. These scholars, Richard

[127] Vines (*Problem*, 63) states: "genre is, strictly speaking, not about formal similarities, but about ideological trajectories, and the best indicator of these trajectories in the chronotope."

[128] Aune, "Genre Theory," 146 n 46.

[129] Thus, Smith (*Why βίος*, passim) may be a bit premature in assuming that the Gospels are biographies and using that as evidence for his case regarding wide readership for the Gospels.

A. Burridge and David L. Mealand, would merit such scrutiny in their own right because they attempt to bring quantitative, testable criteria to what has traditionally been a very impressionistic endeavor.[130] Furthermore, given that I am going to be using a quantitative methodology myself, I want to define clearly where I can perhaps move the discussion forward. Before discussing next steps, however, I turn to discussing the steps they have taken, beginning with Burridge.

Several recent writers on the Gospel genre call attention to the pivotal importance of Burridge's work. Justin M. Smith credits Burridge with having "made remarkable strides in turning the scholarly consensus toward viewing the Gospels as examples of βίοι."[131] As mentioned above, Smith himself acts as though the case is closed, proceeding from the assumption that the Gospels are in fact biographies, and moves on to the further question of how this might impact our understanding of the audience(s) for whom the Gospels were written.[132] Similarly, Andrew Pitts, whose dissertation is largely devoted to critiquing Burridge's work, recognizes that it "continues to function as the foundation for much Gospels research" because it assembles a number of formal criteria for potentially detecting similarities between the Gospels and Greco-Roman βίοι.[133]

Four sorts of features provide these criteria: opening features (title, opening words, preface, etc.), subject, external features (e.g. order of presentation, meter, length, structure), and internal features (e.g. tone, attitude, quality of characterization).[134]

---

[130] On a related note, Jockers (*Text Analysis*, vii–viii) cogently argues for the value in collecting quantitative data on a question like this, even if these data wind up supporting the status quo belief reached on impressionistic grounds.

[131] Smith, *Why βίος*, 36; cf. Walton, "Burridge's Impact," 81–93.

[132] Smith, *Why βίος*, passim; cf. Burridge, "About People," 113–45; Walton, "Burridge's Impact," 87–88.

[133] Pitts, "Genre," 51–52.

[134] Burridge, *Gospels*, 107.

Particularly interesting in light of the direction I want to take my work is Burridge's attempt to describe the subject of a work by quantifying the distribution of nominative nouns within it.[135] Bringing hard linguistic data into play is what makes Burridge's work more persuasive than previous attempts to compare the Gospels with Greco-Roman biographies.[136]

However, despite the influential and suggestive nature of Burridge's work, Pitts raises several cogent objections. The crux of these is that Burridge's net will catch far more than the biographies in which he is interested.[137] Pitts devotes the majority of his attention to so-called "disambiguation criteria," i.e. criteria that will differentiate between biography and related genres, which the majority of Burridge's criteria cannot do effectively.[138] In fact, Pitts points out, a number of them are flawed even as detection criteria.[139]

Pitts finds only one of Burridge's criteria to be potentially effective as a disambiguation criterion, namely the distribution of nominative nouns, but he finds four hurdles to using it successfully in such a way.[140] First, Burridge's comparative corpus contains too little diversity in terms of genre to substantiate that proper nouns in the nominative case occur more densely in biography.[141] Second, Pitts criticizes "Burridge's lack of comprehensiveness in his treatment of Greek subjects": there is far more to the subjects of Greek verbs than simply proper nouns in the nominative case.[142] Third,

[135] Burridge, *Gospels*, 110–12.
[136] Cf. Güttgemanns, *Candid Questions*, 3.
[137] E.g. Pitts, "Genre," 21.
[138] Pitts, "Genre," 63–124.
[139] Pitts, "Genre," 36–39.
[140] Pitts, "Genre," 36.
[141] Pitts, "Genre," 41–42.
[142] Pitts, "Genre," 43.

Burridge does not develop a linguistic framework that would allow him to link grammatical subjects to the notion of discourse topic, and he neglects the distinction between obligatory and non-obligatory subjects.[143] Lastly, Burridge fails to account for the effect of language formality on sentence structure: by tying his count of nominative subjects to the number of sentences in the discourse Burridge overemphasizes works characterized by coordination instead of subordination.[144]

Pitts applies Burridge's nominative distribution criterion to Appian's *Civil War,* which is clearly a history, but Burridge's criterion would designate it a biography.[145] In fairness, Burridge realized that this criterion was "a rather blunt object."[146] The hurdles that Pitts identifies all result from the same underlying cause: Burridge has failed to integrate the resources that modern linguistic models could have provided him.[147]

David L. Mealand's work on "stylometry" also attempts to bring concrete linguistic data to bear on the nature of the Gospel genre, among other issues.[148] Mealand's programmatic article calls attention to several different stylometric tests, including measuring sentence length, the frequency of καί, the place various conjunctions occur in a sentence ("positional stylometry"), the frequency with which a particular part of speech occurs as the last word of a sentence, proportions of various tense-form/mood combinations or case usage ("grammatical stylometry"), and even syntactical patterns

---

[143] Pitts, "Genre," 44–45.
[144] Pitts, "Genre," 46–47.
[145] Pitts, "Genre," 49–51.
[146] Burridge, *Gospels*, 111.
[147] Hägg (*Art of Ancient Biography*, 114) points out that Burridge's study is "based on a rather limited selection of texts." This is another area in which modern linguistic models can be of use.
[148] Mealand, "Computers," 97–115; Mealand, "Hellenistic Historians," 42–66; Mealand, "Hellenistic Greek," 323–45.

("syntactical stylometry").[149] Two of his later articles apply this model to compare the styles of Luke–Acts to the styles of the various historians of the Hellenistic age.

The first article, published in 1991, documented substantial parallels in vocabulary and syntactical usage between Luke–Acts and Hellenistic historians, especially Polybius.[150] Though these data are potentially suggestive of a generic connection, they are too limited in scope to be definitive.

Mealand's 2012 article "Hellenistic Greek and the New Testament: A Stylometric Analysis" further builds his case.[151] In this article Mealand uses Correspondence Analysis, a complex multivariate statistical technique, to generate a plot diagram that ranks samples from the Septuagint, the New Testament, and contemporary Hellenistic authors based on how they use certain conjunctions and morphemes.[152] The horizontal dimension roughly corresponds to the linguistic level, and the vertical axis seems to represent a genre difference in that narrative texts score far lower than argumentative works like epistles and technical treatises.[153]

Mealand should be applauded for his attempt to bring hard data to bear on this question, and he demonstrates substantial similarities between literary Hellenistic historians like Polybius and Luke–Acts with regard to both linguistic level and genre (if one grants that the vertical axis measures genre).[154] Mealand also does a commendably thorough job with his statistical analysis. However, statistics are just a tool for analyzing data; such tools do little to inform the researcher whether he or she has collected the

[149] Mealand, "Computers," 107–8.
[150] Mealand, "Hellenistic Historians," 46–63.
[151] Mealand, "Hellenistic Greek," 323–45.
[152] Mealand, "Hellenistic Greek," 327.
[153] Mealand, "Hellenistic Greek," 328–29.
[154] Mealand, "Hellenistic Greek," 330, 333.

relevant data in the first place. As with computers, statistical measures cannot rise above the quality of the input they are given—garbage in, garbage out.[155] Characterizing Mealand's data as garbage would be unduly harsh, but—however politely one chooses to phrase it—the linguistic features Mealand studied were too few and "idiosyncratic," rendering the results of the statistical analysis unreliable.[156]

Mealand apparently assumes that a straightforward relationship between lexical distribution and genre is self-evident. That might be true to some extent for content words, since similar content is one of the criteria for assigning a group of works to the same genre. Unfortunately, Mealand explicitly prefers function words.[157] The best one can say for focusing on function words is they apparently "exsanguinate our hapless corpus the *least*," that is, they accidentally represent a balanced language sample.[158] Additionally, genre assignment depends on both similar form and content. Vocabulary statistics shed no light on issues of form because units of discourse above the clause—let alone the word—are involved.

The attempts that Burridge, Mealand, and others have made to bring concrete linguistic data to bear on the question represent a solid step forward. Such hard data are intersubjectively verifiable and thus seem more likely to lead to consensus than subjective, impressionistic methods. However, as Pitts and Libby point out, serious methodological questions remain unanswered in that Burridge and Mealand muster no convincing model for how the frequency of concrete forms present on the surface of a text or texts (e.g. proper nouns in the nominative case or certain particles) affect abstract

[155] Cf. Stubbs, "On Texts," 129.
[156] Libby, "Disentangling," 146–48.
[157] Mealand, "Hellenistic Greek," 326.
[158] Libby, "Disentangling," 104.

features (e.g. genre).[159] There is no convincing basis for comparison between texts without such a model.[160] In essence, previous quantitative efforts to study this question have run afoul of what John Sinclair refers to as "a professional hazard" for practitioners of quantitative linguistic methods: such scholars "are always at risk of having their scholarly reports greeted with cries of 'So what?'"[161] Chapters 2 and 3 offer a solution to this methodological lacuna. Before I pursue this thread farther, however, there remain some issues to cover in this survey of previous scholarship. In particular, I now turn to characterizing the Greek of the New Testament.

How Close is Mark to the Attic? The Language of the New Testament

Questions regarding what sort of Greek the New Testament contains arose at several points over the course of the previous section. By way of reminder, the form-critical consensus in favor of the Gospels as *sui generis* pieces of literature rested on the judgment that the Gospels were not literary products and, therefore, comparing them to recognized literary genres in the contemporary milieu was a waste of time. I noted above that—even granting them their premise—this argument does not follow logically; here I further argue that their premise was not true, either. Furthermore, one of the assumptions underlying Tolbert's assessment of the Gospels and Greco-Roman novels appears to be that both are written in simple, unliterary style. Thus, assessing the literary level of the Greek of the New Testament—and of the Gospel of Mark in particular—is a vital part of evaluating previous proposals regarding Mark's genre.

---

[159] Libby, "Disentangling," 32–33; Cf. Halliday, "Machine Translation," 27.
[160] Cf. Halliday, *Grammar*, xxii.
[161] Sinclair, "Introduction," 1.

The nature of the language of the New Testament has been a controversial one.[162] This question has a history going back to at least the fifth century.[163] Proposals in the modern era, over the last 150 years or so, have fallen at all points across the spectrum from "a pure koine derived directly from Attic Greek to a heavily Semitized translation Greek."[164] Between these two extremes, a third popular option has been to argue for a "Jewish-Greek dialect in use in Palestine in the first century."[165]

Three subsections below survey the three major positions, beginning with the one I think is least likely and moving upward from there. Each survey briefly describes the essence of the position, some of the major figures in the debate who held the position, and lastly how discussion of Mark in particular has affected the discussion of the perspective in question. The fourth and final subsection evaluates the three perspectives vis-à-vis each other, concluding with a discussion of how this evaluation impacts the question of the nature of the gospel genre.

Translation Greek?

Although hard evidence is lacking, the idea that parts of the New Testament originated in a Semitic language has a long history, arguably going back to the early second century, depending on how one understands Papias' testimony regarding Matthew. Scholars have typically recognized Aramaic as the Semitic language in question, but a few have argued for Hebrew.[166] Porter indicates that this theory outstrips other challenges to the Deissmann-Moulton theory, covered below, in terms of both number of supporters and

---

[162] Cf., e.g. Porter, "Greek of the New Testament," 11–38; Watt, "Brief History," 237–41.
[163] Maloney, *Semitic Interference,* 1.
[164] Porter, "Greek of the New Testament," 12.
[165] Porter, "Greek of the New Testament," 28.
[166] Porter, "Greek of the New Testament," 22–24.

the strength of the rhetoric with which they defend it.[167] Porter mentions several stages through which this perspective has passed: an early stage (e.g. Blass, Dalman, Wellhausen), a middle stage (e.g. Torrey, Burney, J. A. Montgomery), and a late stage (Black, Fitzmyer).[168]

In terms of the early stage, Gustav Dalman was primarily known for generating the scholarly consensus, which has largely endured, that Aramaic was the primary language of first-century Palestine.[169] He attempted to differentiate between two spoken dialects of Aramaic distributed along geographical lines, Galilean and Judean.[170] He thought he could isolate these geographical dialects in rabbinic literature, especially the Jerusalem Talmud and various midrashim.[171] Alongside these, according to Dalman, was a "uniform literary" variety, available to modern scholars through the Onquelos Targum; this is the variety he used as his control to document Aramaisms in the Synoptic Gospels.[172]

Maloney points out that the first edition of Wellhausen's *Einleitung in die drei ersten Evangelien* claimed "any non-Greek expression could be a Hebraism or an Aramaism," turning a blind eye to contemporary work on the papyri in defining what qualified as "non-Greek," but the second edition "made a concession to contemporary Koine scholarship" seemingly by taking a position more similar to Gehman and Turner (see below).[173] Maloney further notes that Wellhausen went far out on the limb of

---

167 Porter, "Greek of the New Testament," 17–18.
168 Porter, "Greek of the New Testament," 18–22.
169 Maloney, *Semitic Interference*, 11.
170 Maloney, *Semitic Interference*, 11.
171 Maloney, *Semitic Interference*, 12.
172 Maloney, *Semitic Interference*, 12.
173 Maloney, *Semitic Interference*, 11.

Semitic influence in his commentaries, "pointing out Semitisms and using almost any Semitic parallel he could find as proof."[174]

The middle stage, found in the work of Torrey, Burney, J. A. Montgomery, and M. Burrows, "explored further the evidence for syntactic Semitisms" in the service of a translation theory to explain the nature of the Gospels and Acts.[175] Torrey was critical of Dalman's choice of control literature in principle, but Maloney points out that in practice he used whatever source helped his case, including Syriac.[176] Torrey denied biblicisms, preferring to attribute these directly to Aramaic.[177] Burney cast his net almost as wide as did Torrey.[178] Montgomery discussed all of the possibly relevant Semitic languages (Hebrew, Aramaic, and Syriac), but "omitted discussion of the Greek evidence in any detail, and did not even give specific parallels for the supposed Semitic equivalents."[179] M. Burrows considered all of Mark to be translation Greek because he could not find seams by looking for "'Aramaic coloring'".[180]

"Thy Speech Betrayeth Thee": A Special Jewish-Greek Dialect?

Another major perspective, in many ways intermediate between understanding the Greek of the New Testament as overliteral translation from a Semitic *Vorlage*, discussed above, and the idea that it is a more or less idiomatic part of Hellenistic Greek as a whole, is the idea that a specifically Jewish dialect of Greek existed during the first century.[181] Modern

[174] Maloney, *Semitic Interference*, 11.
[175] Maloney, *Semitic Interference*, 12.
[176] Maloney, *Semitic Interference*, 12.
[177] Maloney, *Semitic Interference*, 12.
[178] Maloney, *Semitic Interference*, 13, citing Burney, *Aramaic Origin.*
[179] Maloney, *Semitic Interference*, 13, citing Montgomery, *Origin of the Gospel of St. John*.
[180] Maloney, *Semitic Interference*, 13, citing Burrows, "Transitions," 117–23.
[181] Porter, "Greek of the New Testament," 27–31; cf. Watt, "Brief History," 237–38.

proponents of this theory include Henry Gehman and Nigel Turner, but its roots go back to the idea of "Holy Ghost Greek" advocated by scholars of the nineteenth century.[182]

Porter notes that "Gehman hides nothing with respect to the basis for his position, going into detailed discussion of the examples he uses."[183] He also notes that one factor differentiating the positions of Gehman and Turner is that the former believed that the distinguishable Jewish dialect "may have been a temporary linguistic condition" as Jews moved from a purely Semitic environment to interaction with Hellenistic environment, while the latter conceived of the distinctive Jewish dialect as a permanent, living variety.[184]

Nigel Turner argues that Mark—particularly the form found in Codex Bezae—is the "most characteristic form" of his supposed Jewish Greek dialect.[185] In another work, his stated purpose is "to expose consistently the almost complete absence of classical standards in early every author."[186] Later in the same discussion he suggests that the standards of Biblical Greek potentially come from "outside secular Greek altogether," although he admits that "the living Koine must be kept in mind always."[187] In another place he refers to "a family likeness among these Biblical works, setting them apart from the papyri and from contemporary literary Greek."[188] For Turner, the setting apart of Biblical Greek is separate from issues of the linguistic background of the author and his audience: "It does not follow that if a construction occurs as frequently in the epistles as

[182] Porter, "Greek of the New Testament," 27–28.
[183] Porter, "Greek of the New Testament," 30.
[184] Porter, "Greek of the New Testament," 29.
[185] Turner, *Style*, 7.
[186] Turner, *Syntax*, 2.
[187] Turner, *Syntax*, 3.
[188] Turner, *Syntax*, 5.

in the gospels it will be less likely to have a Semitic origin" because the Septuagint's "idioms powerfully influenced free compositions of Biblical Greek."[189]

These scholars have not successfully responded to Deissmann's argument that similarities of phonology and morphology leave the theory of special Jewish Greek "shattered beyond repair."[190]

As a parallel to Turner, Porter quotes from Matthew Black's article on the biblical languages in the *Cambridge History of the Bible*: "Black has even gone so far as to call biblical Greek 'a peculiar language, the language of a peculiar people.'"[191] The effect this quotation had on me aptly summarizes how the Gehman-Turner position sees the language of the New Testament. For me, as someone raised in an environment within which the KJV was nearly unchallenged as the version of choice, just as the LXX would have within the environment Gehman and Turner posit, Black's statement has inescapable value judgment associated with it, for "peculiar people" is how the KJV renders the phrase that most modern translations render "a people for God's own possession," or something to that effect, in 1 Pet. 2:9. Noticing this allusion leads me to raise two points by way of closing this sub-section. First, this allusion clearly demonstrates the idea of "Holy Ghost Greek": the special people, the "people for God's own possession," use a special language—presumably also God-given. Second, despite being fairly certain no stranger who met me on the street would suspect me of speaking some special seventeenth-century dialect of English, I recognized Black's wording immediately and have been known to throw "three-score and ten" into a conversation—or

[189] Turner, *Syntax*, 4.
[190] Deissmann, "Hellenistic," 51.
[191] Porter, "Greek of the New Testament," 29; cf. Deissmann, "Hellenistic," 50.

to name a sub-section "Thy Speech Betrayeth Thee" (Matt 26:73, KJV) for that matter.[192] Perhaps, by analogy, the influence of the LXX and religious usage is too slender of a reed to support the hypothesis of a separate Jewish Greek dialect.

Knowledgeably Keeping Koine

The final category of opinions I will cover on the relationship between the Greek of the New Testament and other Hellenistic Greek is the perspective that they are very similar, if not identical. A number of scholars have taken positions along this line, famously including Adolf Deissmann and James Hope Moulton around the turn of the twentieth century.[193] More recently, Lars Rydbeck and Moises Silva have proffered similar answers.[194]

One recent investigation into the role of Semitic syntax in the New Testament, and Mark in particular, recognizes the work of Deissmann as having "launched the modern period of New Testament grammar."[195] Adolf Deissmann published extensively on the nature of Hellenistic Greek, including an article for a major German dictionary.[196] Deissmann defines Hellenistic Greek as all Greek between 300 BCE and 600 CE, noting that neither defining it as the language of Jews and Christians or as that which falls short of Attic style is helpful.[197] Deissmann also calls attention to the need to separate multiple linguistic levels within Hellenistic Greek, noting that the "literary artifacts" of the Hellenistic age often, though not always, betray a tendency to emulate the great Attic writers of the past over against what seems to be the tendency of the living language,

---

[192] Cf. Moulton, "New Testament Greek," 67.
[193] Porter, "Greek of the New Testament," 12–17.
[194] Porter, "Greek of the New Testament," 31–33.
[195] Maloney, *Semitic Interference,* 7.
[196] E.g. Deissmann, "Hellenistic," 39–59.
[197] Deissmann, "Hellenistic," 41.

judging from the papyri, inscriptions, etc.[198] This is true even within the New Testament itself.[199]

Maloney calls attention to the work of A. Thumb in supporting Deissmann.[200] He went further than Deissmann, not only recognizing that the New Testament was not the product of a special Jewish dialect, but also showing a direct relationship between the vernacular Koine and modern Greek, so that we have both a beginning point (Attic) and an endpoint (Modern Greek) within which we can find trajectories to explain potential Semitisms.[201]

It fell to James Hope Moulton to expand Deissmann's research, which was largely lexical, into grammar and syntax. Similar to the way Deissmann was able to parallel a number of unusual lexical items in the New Testament from the papyri, Moulton did so with grammatical and syntactical items, including the instrumental use of ἐν, which former scholars had thought a Semitism.[202] Particularly apposite in this regard, given the counter-argument of Semitic influence on the papyri, is Moulton's argument relating the provenance of the parallel to the probability of Jewish occupation: "Then in 1902 appeared the first volume of papyri from Tebtunis, with half-a-dozen examples of ἐν μαχαίρῃ and the like, all due to different writers…Are we to explain the new 'Semitisms'

---

[198] Deissmann, "Hellenistic," 44.

[199] Deissmann, "Hellenistic," 58.

[200] Maloney, *Semitic Interference,* 8; cf., e.g., Thumb, *Sprache,* passim. No less a scholar than James Hope Moulton ("Thumb," 221–22) calls attention to the wide range and influence of Thumb's work in an obituary, stating "Thumb's range of learning and research was immense, covering well-nigh everything that came under the comprehensive heading Indo-European linguistic" (221); his published work on Greek ranged from "prehistoric foundations" to "the patois of the modern peasant," applying the latter to the knowledge of the κοινή (221).

[201] Maloney, *Semitic Interference,* 8.

[202] Moulton, "New Testament Greek," 65.

by postulating an influential Jewish colony at or near Tebtunis—the seat, by the way, of a 'famous' (λόγιμον) temple of the crocodile-god Sobk?"[203]

E. J. Goodspeed, according to Maloney, "pressed the point of papyrus parallels almost to the extreme of Deissmann and the early Moulton, claiming that Aramaic originals would be a priori impossible since no Aramaic literary tradition had been found to date."[204] As an example of Goodspeed's extreme tendency towards papyrus parallels, to which Maloney refers, consider his comparing of the word ῥακά in Matthew 5:22 to a papyrus fragment of the third century B.C.[205] Significantly, Goodspeed was writing before the discovery of the Dead Sea Scrolls. Nevertheless, he makes a solid case against an Aramaic original for Acts, pointing to the opening and closing of the letter emerging from the Jerusalem council (Acts 15:23–29) as "in epistolary forms the most perfectly Greek letter in the New Testament," with the second like unto it also being in Acts (23:26–32).[206] Another of Goodspeed's works chastises proponents of the Aramaic hypothesis for failing "to exhaust all available lexical and literary aids in that particular tongue" in which the NT is extant "before exploring remote corners of other languages for light."[207]

G. D. Kilpatrick's work "Some Notes on Marcan Usage" includes several notes that are apposite in this area.[208] He quite correctly calls attention to Atticizing tendencies on the part of the scribes as a complicating factor in determining how Mark's usage

---

[203] Moulton, "New Testament Greek," 65–66.
[204] Maloney, *Semitic Interference*, 14; cf. Goodspeed, "Origin," 87–89; Goodspeed, *New Chapters*, 127–68.
[205] Goodspeed, *New Chapters*, 147–48.
[206] Goodspeed, "Origin," 87.
[207] Goodspeed, *New Chapters*, 136.
[208] Kilpatrick, "Some Notes," 159–74.

relates to Attic.[209] Also interesting is his positioning Mark in the diachronic development of the use of the so-called "resultative perfect": though I find the entire notion of this category dubious, he says this meaning "is predominant" in classical prose and "continues frequent" in the papyri and Atticizing Greek, but in Modern Greek "it survives only where literary influences are strong enough to maintain it."[210] In this respect, Mark falls closer to Modern Greek.[211]

This debate has affected the study of Mark.[212] For instance, Maloney concludes that "syntactical Semitic interference permeates every page of the gospel."[213] Rüger notes a total of twenty-one "lexical Aramaisms," including proper names, "individual words and short sentences."[214] The proper names are self-explanatory, but the latter two categories refer to expressions like ἀββά (Mk 14: 36) or ἐλωῒ ἐλωῒ λεμὰ σαβαχθάνι (Mk 15:34), respectively.[215]

Summary

What Porter and O'Donnell have to say about the study of conjunctions is much more broadly applicable: "Even the work of those who are linguistically informed often suffers from approaching the Greek of the New Testament either atomistically or in terms of Semitic influence—two approaches that inevitably skew the evidence and understanding."[216] Decker concludes that Mark's Greek "is toward the less literary end of

---

[209] E.g. Kilpatrick, "Some Notes," 160. For a description of Atticism and its relevance to the NT, see the work of Sean A. Adams ("Atticism," 93–100) and John A. L. Lee ("Atticist Grammarians," 283–308).
[210] Kilpatrick, "Some Notes," 164.
[211] Kilpatrick, "Some Notes," 164.
[212] E.g. Decker, "Markan Idiolect," 63–65.
[213] Maloney, *Semitic Interference,* 245.
[214] Rüger, "Die lexikalischen Aramaismen," 73.
[215] Rüger, "Die lexikalischen Aramaismen," 73, 78.
[216] Porter and O'Donnell, "Conjunctions," 5.

the spectrum," but this "ought not be over-emphasized since all Koine would be judged poorly if compared with the prose of the Attic luminaries."[217]

One of the key issues in evaluating this debate is recognizing that the various scholars involved have not always defined their terms clearly.[218] For example, Nigel Turner states: "By Aramaisms, Hebraisms and Semitisms respectively, are intended those Greek idioms which owe their form or the frequency of their occurrence to Aramaic, Hebrew, or an influence which might equally well apply to both languages."[219] Furthermore, when comparing various languages, scholars need to understand clearly when the filter of their mother tongue is confusing their analysis. For example, C. H. Turner defines the "impersonal plural" as being "the use of a plural verb with no subject expressed, and no subject implied other than the quite general one 'people'."[220] This manner of expression, "common in Aramaic as a substitute for the passive" supposedly "is very characteristic of St. Mark," but the other Synoptic evangelists typically change it.[221] Beyond differences of opinion on whether all of the examples Turner gives of the impersonal plural even meet his own stated definition, I would like to raise a larger issue: has not the filter of English, which requires an explicit subject, affected how he sees the relationship between Aramaic and Mark's Greek?

---

[217] Decker, "Markan Idiolect," 64–65.
[218] Porter, "Greek of the New Testament," 35–36.
[219] Turner, *Style*, 5.
[220] Turner, "Marcan Usage," 4.
[221] Turner, "Marcan Usage," 4.

Clause Component Ordering in Hellenistic Greek

Background to Clause Component Ordering Debates

The body of this project compares Mark to two candidate analogical genres, which I call referential and non-referential narrative, using statistical significance tests to measure whether random chance can sufficiently explain the differences observed in patterns of clause thematization. Chapters 2 and 3 discuss the details of this process, but suffice it to say for the time being that my project is concerned with the resources Koine Greek provides authors of texts for organizing the clauses of which the text consists, i.e. what linguists call information structure.[222] As Runge points out, NT studies has traditionally addressed issues of information structure under the rubric of word order.[223] Therefore, the history of discussion on word order in the Greek of the New Testament is relevant background to my project.

As recently as 2015—a little over twenty years after his first article on the subject—Stanley E. Porter re-raised an important question relative to my study, namely "[i]s Greek word order still an unexplored area in New Testament studies," concluding that "[t]he brief answer is yes."[224] Porter's review of an extensive body of literature reveals the hyperbolic nature of the essay's title.[225] Nevertheless, it is fair to conclude that the study of word order in Hellenistic Greek is an *under*explored area and that much of what has been done has been a vague combination of impressionistic judgments and very rudimentary attempts at quantitative methodology—reminiscent of Burridge and

---

[222] Cf., e.g., Runge, *Discourse Grammar*, 181–205.
[223] Runge, *Discourse Grammar*, 181.
[224] Porter, "Greek Word Order—Still Unexplored," 347; cf. Porter, "Word Order and Clause Structure," 177–204.
[225] Porter, "Greek Word Order—Still Unexplored," 348–53.

Mealand, in fact. To borrow an image from the book of Judges: there has been no rigorous method in the land, so everyone's conclusions are those that are right in his or her eyes, i.e. that his or her argument requires. For example, C. H. Turner's "Notes on Markan Usage," originally published between 1924 and 1928, argue that Mark pushes the verb to the end of the sentence, while Nigel Turner's *Syntax* (1963) argues that Mark normally places the verb as close to the beginning of the clause as possible.[226] This disagreement is particularly egregious, but the difference is more of degree than kind. Virtually the only point on which all parties agree is that word order in Hellenistic Greek is freer than in English.[227]

In fairness, this situation began to change during the last twenty years, largely due to the work of Porter himself and some of his former students.[228] The burden of the recent essay is to clarify three areas of confusion that have stymied work in this area.[229] First, a study in this area should carefully delineate "what level of study (place on the rank scale) it is interested in analyzing."[230] Second, a study in this area should make sure to clearly distinguish between form and function.[231] Third, however, the "most pressing need is persuasive explanations of the patterns observed."[232]

---

[226] Turner, "Notes on Markan Usage," 126–30; Turner, *Syntax*, 170.

[227] E.g. BDF §472; cf. Porter, "Word Order and Clause Structure," 177–78.

[228] Some of these are: Porter and O'Donnell ("Conjunctions," 3–14 and "Vocative," 35–48), Pitts ("Word Order," 311–346).

[229] Porter, "Greek Word Order—Still Unexplored," 353–61.

[230] Porter, "Greek Word Order—Still Unexplored," 354. Pitts ("Word Order," 321–22) details how failing to abide by this admonition has led previous studies astray. The term "rank scale" comes from the work of the linguist M. A. K. Halliday, to whom I will devote a great deal of attention in Chapter 2.

[231] Porter, "Greek Word Order—Still Unexplored," 355–57.

[232] Porter, "Greek Word Order—Still Unexplored," 357.

Previous Work on Word Order in Classical Greek

I will briefly summarize some of the work done previously on word order and clause constituent order in classical Greek. This work is important to include here because it has influenced work on Hellenistic Greek.[233]

The first study to discuss is Kenneth J. Dover's book, *Greek Word Order*. Writing in 1960, Dover observes that "[t]he problem of Greek word order is so seldom discussed in this country that it is still possible to treat it as a fresh problem. In this respect it differs from all other problems of comparable magnitude in the study of the Greek language," and he characterized the limited amount of work that had been done as mostly "cautious and limited in scope."[234] This focus on small-scale details arose, in Dover's opinion, because the rules suggested in more ambitious studies looking for consistently applicable patterns either "break down" upon attempting to apply them to a randomly-selected page of Greek text or are ultimately circular and "admit neither of proof or disproof."[235] He humorously summarizes the resulting attitude of just stating the facts, rather than attempting to explain them: "it amounts to saying 'xyz and xzy occur, but, on the other hand, yxz, zyx, zxy, and zyx also occur.'"[236]

Dover himself takes a middle course between the Scylla of rigid, invariant rules and the Charybdis of eschewing any sort of explanation, noting both "processes in the mind of the composer, rational or irrational, which we cannot necessarily expect to

[233] E.g. Pitts, "Word Order," 311–12. The reverse is not true: Dover (*Word Order*, 10) mentions the Gospels and Plutarch as potential sources, but only to dismiss their value when classical sources are available.
[234] Dover, *Word Order*, 1.
[235] Dover, *Word Order*, 1.
[236] Dover, *Word Order*, 1.

recover" and that a "limited number of patterns" account for much of the variation.[237] Dover's procedure for explaining word order variation consists of six steps: (1) picking a particular candidate cause to investigate, (2) intuitively constructing rules that are as general as possible while still being falsifiable, (3) gathering texts to investigate, ensuring that the chosen texts do not favor the intuitively constructed rules but not necessarily ensuring that the texts are representative of an entire language variety, (4) count how often the selected texts follow the intuitively constructed rules versus how often the rules do not account for the observations, (5) distinguish between normal and abnormal patterns on the basis of these counts, and (6) attempt to explain the exceptions by finding things that are true of all the exceptions but not true of any cases that follow the intuitively constructed rules.[238] His results indicate that subjects usually precede the verb, but the ratio of pre-verbal objects to post-verbal ones "shows greater fluctuation, the extremes being 0.7 and 4.0, but in most types of clause in most authors it exceeds 1.0."[239] A ratio of 1.0 would indicate that the two were equally frequent, so Dover's results point to objects being more frequent at the beginning of a clause than at the end. Unfortunately, Dover does not consider equative clauses and lumps direct and indirect objects together.[240] As a result, his findings are potentially skewed, and (even if accurate) he is forced to conclude that his "statistics are very far indeed from establishing for 'Classical Greek' *simpliciter* anything worth calling a syntactical rule of word order."[241]

---

[237] Dover, *Word Order*, 2. Significantly, given the direction my research is heading, Dover (*Word Order*, 4) suggests social considerations as one factor that could account for word variation.

[238] Dover, *Word Order*, 4–5.

[239] Dover, *Word Order*, 25.

[240] Dover, *Word Order*, 26.

[241] Dover, *Word Order*, 31.

A more recent article by Graham Dunn argues "that although Dover's claim is true in some respects it is misleading with regard to the general picture which has emerged from our research."[242] Dunn recognizes that word order in Greek is a matter of statistical tendency, rather than invariant laws.[243] Dunn draws his data from Book I of Herodotus.[244] The theoretical model is a dependency one, dividing "every Greek construction consisting of two or more words…into two parts, the head and the modifier"; specifically, for Dunn's study, "main verbs" are considered head terms and other elements are considered modifiers.[245] His procedure consists of counting the number of times the head preceded the modifier and the number of times the modifier preceded the head in several different environments and then performing a χ^2 analysis using these two counts as the observed frequencies and the average of the two counts as the expected frequency with a significance level of 5 percent.[246]

The environments for which Dunn collects separate frequency counts are "subordinate clauses," "participial clauses," "noun phrases," and "prepositional phrases."[247] Dunn identifies statistically significant tendencies for all the categories into which divides subordinate clauses: temporal and conditional clauses tend to precede the verb, while the others tend to follow it.[248] Turning to participial clauses, the genitive absolute, the dative participle, aorist participles in the nominative case tend to come before the verb; present participle in the nominative case, perfect participle in the

[242] Dunn, "Syntactic Word Order," 78.
[243] Dunn, "Syntactic Word Order," 63.
[244] Dunn, "Syntactic Word Order," 64.
[245] Dunn, "Syntactic Word Order," 64.
[246] Dunn, "Syntactic Word Order," 64–66.
[247] Dunn, "Syntactic Word Order," 66. Dunn ("Syntactic Word Order," 66) clarifies that infinitival clauses fall under subordinate clauses, and the "manner adverbial" is treated as a noun phrase.
[248] Dunn, "Syntactic Word Order," 69. A minus signifies a significant tendency precedes the verb; a plus signifies a significant tendency to follow the verb (Dunn, "Syntactic Word Order," 67).

nominative case, the future participle in the nominative case, accusative participles functioning as objects, and predicate nominative participles follow the verb.[249] With noun phrases, Dunn was unable to identify statistically significant norms for the placement of direct objects, "subjectival objects," and instrumentals.[250] The equational complement, a term which Dunn does not define but I presume to be a noun functioning as a predicate nominative, tends to follow the verb.[251] Vocatives, temporal phrases, subjects, "passive subject[s]," which I take to be the subjects of intransitive verbs, "equational subject[s]," and manner adverbials tend to precede the verb.[252] Dunn found a statistical norm for only one category of prepositional phrase: as one would expect prepositions indicating the destination of an action tend to follow the verb.[253] On the basis of these data, Dunn infers that "the Greek sentence emerges as verbicentric, i.e. having the verb at the centre with modifiers on either side."[254]

Pre-Linguistic Studies of New Testament Clause Structure

Discussion of word order and clause structure has a long history in studies of the Greek of New Testament.[255] Typically, such investigation has tried to prescribe rules for the order of clause constituents. No consensus has emerged from this process, however. The literature contains a morass of mutually contradictory claims regarding the supposed

249 Dunn, "Syntactic Word Order," 72.

250 Dunn, "Syntactic Word Order," 75. Dunn uses the term "subjectival object" to refer to accusatives that function as the subject of an infinitive (Dunn, "Syntactic Word Order," 74n13).

251 Dunn, "Syntactic Word Order," 75.

252 Dunn, "Syntactic Word Order," 75.

253 Dunn, "Syntactic Word Order," 77.

254 Dunn, "Syntactic Word Order," 78.

255 The survey of literature below draws heavily on Porter's ("Greek Word Order—Still Unexplored," 348–53). Porter ("Greek Word Order—Still Unexplored," 348n3) in turn, notes his dependence on Kwong (*Word Order*, 3–29). However, I have also added mention of some additional works, including grammars, which Porter explicitly leaves outside his scope, and some works that are of particular relevance to Mark (e.g. Turner, "Notes on Markan Usage," passim).

normal order of clause constituents for the NT as a whole, particular genres within it, or even a single book like Mark. This survey below begins with treatments of the New Testament, then treatments of narrative, and finally Mark of itself.

One of the most common reference grammars in New Testament studies asserts that, despite Greek word order being "freer by far than in modern languages," the NT (and, particularly, the narrative portions thereof) have "something like a normal word order," namely coordinating conjunction, verb followed by "subject, object, supplementary participle, etc."[256] This tendency is said to be especially true for Mark.[257]

The work of A. T. Robertson follows the same pattern.[258] Robertson stresses how free Greek is from "artificial rules" in the area of word order, attributing it to Greek's rich morphology.[259] According to him, the "only unalterable rule in the Greek sentence" is "spontaneity."[260] Nevertheless, he suggests emphasis as "one of the ruling ideas in the order of words," recognizing that the author creates this emphasis by putting a word in an unusual location.[261] This recognition, in turn, presupposes some notion of usual order such that deviations from it can produce a pragmatic effect. At the level of the word group, though Robertson does not use this terminology, the normal order in question is head term and then modifier.[262] At the level of the clause, the "predicate very commonly comes first, simply because, as a rule, the predicate is the most important thing in the sentence."[263] Robertson's examples here show that his "predicate" category includes

---

[256] BDF §472.
[257] BDF §472(1).
[258] Robertson, *Grammar*, 417–25.
[259] Robertson, *Grammar*, 417.
[260] Robertson, *Grammar*, 417.
[261] Robertson, *Grammar*, 417. Robertson (*Grammar*, 419–23) also discusses rhythm as another.
[262] Robertson, *Grammar*, 418–19.
[263] Robertson, *Grammar*, 417.

verbs as well as what have traditionally been called predicate nominatives.[264] The question arises as to whether noun structures with no explicit verb and finite verbs ought to be considered together as a single category.[265]

C. F. D. Moule devotes a few pages to the issue of word order in his *Idiom-Book of New Testament Greek*.[266] The vast majority of this discussion covers lower levels of discourse, but he introduces the section with a "rough-and-ready rule" for clause structure, namely to "*reverse the English order*, so that the emphatic word comes at or near the beginning of the sentence."[267] This follows along the same tack as BDF and Robertson.

One clear clue that more research into the word order of the New Testament in general, and Mark in particular, is warranted is that the two Turners mentioned in the above discussion of Semitisms can claim completely opposite tendencies for Mark. C. H. Turner's "Notes on Markan Usage," originally published between 1924 and 1928, argue that Mark tends to place "[t]he verb at the end of the sentence."[268] Published in 1963, Nigel Turner's *Syntax* offers an interesting perspective on word order, both within the clause and in terms of clause complexing.[269] The former is significant for my present purposes.[270] He argues that the New Testament documents avoid postpositive conjunctions, such as γάρ, because the Semitic languages do not have such forms.[271] He suggests that verbs in Biblical Greek gravitate as near to the front of the clause, whereas

264 Robertson, *Grammar*, 417.
265 Cf. Porter, *Idioms*, 287.
266 Moule, *Idiom-Book*, 166–70.
267 Moule, *Idiom-Book*, 166, emphasis original.
268 Turner, "Markan Usage," 126–30.
269 Turner, *Syntax*, 344–50.
270 Turner, *Syntax*, 347–30.
271 Turner, *Syntax*, 347.

other writings of the day usually placed in the middle of the clause.[272] However, the evidence he adduces for this is highly suspect: Rife, the authority whom he cites, simply worked his way through the book until he found ten instances of "main declarative clauses" with an explicit subject, complement, and verb, and then analyzed the order in which they appear.[273] Other features of word order that Turner mentions—all of which he ascribes to Semitic influence, as one might expect in light of the earlier discussion of his position relative to the nature of the Greek of the New Testament—are the closeness of the article and the noun serving as its head-term, placement of adjectives after their corresponding nouns, placement of genitive qualifiers immediately after the noun to which they are related, placement of "[u]nemphatic direct or indirect personal pronouns . . . closely after the verb," and prepositional phrases after the noun being modified.[274]

Maloney, following Schwyzer, Zerwick, and a number of statistical studies, suggests that the normal word order for an independent clause in Hellenistic Greek is subject first, then verb.[275] He cites a variety of statistical studies, which have studied a variety of authors ranging from Xenophon to the barely literate writers of the documentary papyri, to support this assertion.[276] Maloney studies the narrative framework, i.e., the parts of the Gospel that are not projected speech, and discovers that "in independent clauses introduced by καί within the narrative part of the gospel, Mark's preference for placing the verb before the subject is striking in comparison to the more normal Greek word order."[277] Maloney concludes that the most reasonable explanation

---

[272] Turner, *Syntax*, 347–48; cf. BDF §472; Kilpatrick, "Some Notes," 174.
[273] Turner, *Syntax*, 347n2, citing Rife, "Mechanics," 250–51.
[274] Turner, *Syntax*, 349–50.
[275] Maloney, *Semitic Interference*, 51.
[276] Maloney, *Semitic Interference*, 197.
[277] Maloney, *Semitic Interference*, 52.

for this is Semitic influence, since Semitic languages tend to place the verb at the beginning of the sentence.[278] There are several potential methodological problems with Maloney's analysis. Most of them are simply those that have bedeviled most other analyses of word order in Hellenistic Greek, which I will cover in due course. However, one I will cover here, as it simply relates to which language should be posited as the source of interference. Even if Maloney's data substantiate Semitic interference into the way Mark structures his independent clauses—a point of which I remain unconvinced—would not Hebrew be more reasonable than Aramaic as the source of interference, perhaps by way of the LXX/OG, given that Maloney focused on clauses that begin with καί?[279] After all, and I admit I am no Aramaic expert, is not the waw conversive construction a feature of Hebrew and not Aramaic? This is especially true in light of Maloney finding that this pattern is "common OG usage," since the OG is for the most part translating Hebrew.[280]

Linguistically Informed Studies of New Testament Clause Structure

As with other aspects of Greek grammatical study, studies of word order and clause component ordering have been slow to make use of models informed by modern linguistic study. This situation has begun to change in recent years. Unfortunately, much of this newer work is scattered in journal articles and book chapters, with only a few monographs devoted to the subject and, as yet, very little effect on grammars and other standard reference works. I see hopeful signs that—at least in some quarters—this situation is changing.

---

[278] Maloney, *Semitic Interference*, 53.
[279] As, e.g., Decker ("Markan Idiolect," 48–49) suggests.
[280] Maloney, *Semitic Interference*, 246. The quote is from the key to Maloney's table (244).

One recent intermediate Greek grammar, *Going Deeper with New Testament Greek: An Intermediate Study of the Grammar and Syntax of the New Testament*, devotes almost three full pages to the issue of word order.[281] The authors of this grammar cite a variety of a linguistically informed work, including Porter's *Idioms*, but not his more recent chapter or Pitts' chapter. Moreover, rather than being mere window dressing for the sake of scholarly completeness, these citations clearly influence the discussion. The authors call attention to the fact that pragmatic considerations often affect an author's choice of how he or she structures a clause—sometimes without conscious effort.[282] The table with which they summarize their discussion points to some of the major issues, but it lacks nuance, seeming to leave the impression that all the possibly emphatic word orders they list are equal.[283] Furthermore, in some cases, they could have chosen examples that would have been clearer: their example of an explicit subject preceding its verb is θεὸν οὐδεὶς ἑώρακεν (John 1:18), which also has a complement in thematic position.[284] Nevertheless, those interested in dissemination of linguistically responsible scholarship should commend Köstenberger and his coauthors for this effort.

Another recent intermediate grammar includes "Word and Clause Order" as a heading within its treatment of prominence in discourse.[285] This discussion focuses on the functional impact of "deviation" from normal word-order patterns, calling it "foregrounded."[286] The most common foregrounding device in clause structure is so-called "left-dislocating," moving the material to the front of the clause.[287] Foregrounding

---

[281] Köstenberger et al., *Going Deeper*, 448–50.
[282] Köstenberger et al., *Going Deeper*, 449.
[283] Köstenberger et al., *Going Deeper*, 450.
[284] Köstenberger et al., *Going Deeper*, 450.
[285] Mathewson and Emig, *Intermediate Greek Grammar*, 13.18–19, locations 8278–8362.
[286] Mathewson and Emig, *Intermediate Greek Grammar*, 13.18, locations 8278–8280.
[287] Mathewson and Emig, *Intermediate Greek Grammar*, 13.19, locations 8324–8362.

here is roughly equivalent to what earlier grammarians called emphasis, but as a more fine-grained category.

Shorter discussions of this topic with explicitly linguistic foundations fall into at least two groups according to precisely what linguistic foundations they use. One group, primarily associated with the Summer Institute of Linguistics, has tended towards analysis of lower-level phenomena in particular books or the like, using an eclectic mixture of linguistic models.[288] The other group, basing their models on systemic-functional linguistics developed by M. A. K. Halliday, have tended towards larger-scale projects, both in the sense of considering levels above the clause and studying multiple books—up to and including the entire New Testament.[289]

One example of the SIL approach is Levinsohn's chapter on rules for constituent order and the use of the article in Philippians.[290] This study draws on Simon Dik's Functional Grammar.[291] Levinsohn sets out five principles that he believes explain clause component ordering in Philippians. The first principle relates to moving things to the front of the clause to mark a change in what the author takes for granted as assumed information.[292] Principle 2 identifies items marked out to be "salient" as those immediately following an item topicalized as per the previous principle.[293] Principle 3 allows for the inclusion of a further pre-verbal component following topicalized and salient items, "provided they are 'given'," i.e. taken as assumed information.[294] Principle 4 states: "When the verb is the most salient constituent of a clause, any non-verbal

---

[288] E.g. Levinsohn, "Constituent Order," 60–74; cf. Porter and Pitts, "NT Language," 236–37.
[289] E.g. Porter, "Greek Word Order—Still Unexplored," 347–62; Pitts, "Word Order," 311–46.
[290] Levinsohn, "Constituent Order," 60–74.
[291] Levinsohn, "Constituent Order," 60–61.
[292] Levinsohn, "Constituent Order," 61–64.
[293] Levinsohn, "Constituent Order," 65–67.
[294] Levinsohn, "Consitutent Order," 67–69.

constituent of the predicate precedes it."[295] The final principle differentiates between topicalized and salient information: topicalized information has the article, while salient information is anarthrous.[296]

Stanley E. Porter's 1993 article "Word Order and Clause Structure in New Testament Greek: An Unexplored Area of Greek Linguistics Using Philippians as a Test Case" is an early example of a linguistic approach using Hallidayan linguistics.[297] Under the heading "Word Order," Porter discusses relationships at the level of the word group, including the placement of adjectives, genitives, the article, and the demonstrative pronoun in relation to that which they are modifying.[298] He concludes that, as a general rule, "the headterm has the distinct tendency to precede its modifier."[299]

Under the heading "Clause Structure," Porter discusses several theoretical issues that are directly related to my study.[300] First, in terms of theory, Porter calls into question the categories, derived from linguistic typology, that have typically been the point of departure or these kinds of studies (subject, object, and verb): many Greek clauses do not grammaticalize all three of these elements, and, by the same token, none of them has to be present in any particular clause.[301] As a result, he argues, "it is fruitless when studying the issue of clause structure to hypothesize about the phantom presence of various syntactical phenomena," and "clausal structure must be formulated upon the basis of the explicit structural elements."[302] Second, Porter points out that the subject, if present, "has

---

[295] Levinsohn, "Constituent Order," 69.
[296] Levinsohn, "Constituent Order," 70–74.
[297] Porter, "Word Order and Clause Structure," 177–206.
[298] Porter, "Word Order and Clause Structure," 181–86.
[299] Porter, "Word Order and Clause Structure," 185.
[300] Porter, "Word Order and Clause Structure," 186–203.
[301] Porter, "Word Order and Clause Structure," 186–87.
[302] Porter, "Word Order and Clause Structure," 187.

a distinct tendency to precede its predicate and/or its complement."[303] Third, he suggests differentiating "between main clauses and various forms of dependent clauses, including not only those with finite verbs but also those with infinitives and participles."[304] Fourth, he suggests concentrating on subjects, objects, and verbs in studying clause structure, whereas adjuncts are, for the most part, relevant only as pointers to the presence of subordinate structures.[305] Fifth, he introduces the concept of unmarked and marked word order from modern linguistics, noting that marked word order carries "an attendant meaning not implied by (or in addition to that of) the unmarked syntax," and frequency of appearance is one of the criteria for differentiating between the two.[306] Lastly, he discusses the notion of topic and comment as a way of discussing this issue without recourse to translation, describing the topic as "the portion of the clause that directs the discourse" and the comment as "the portion of the clause that presents information about the topic or elucidates what is already known."[307] The main point of relevance from the application to Philippians, of which the remainder of the article mostly consists, is that a grammaticalized subject "serves as a topic marker and/or shifter."[308] At the end of the article, Porter states that his preliminary investigations of other passages, including one from Matthew and another from Acts, indicate that this pattern holds for narrative literature as well, and the writers of narrative seem to avail themselves of it somewhat more frequently.[309]

---

[303] Porter, "Word Order and Clause Structure," 188.
[304] Porter, "Word Order and Clause Structure," 189.
[305] Porter, "Word Order and Clause Structure," 189.
[306] Porter, "Word Order and Clause Structure," 190.
[307] Porter, "Word Order and Clause Structure," 191.
[308] Porter, "Word Order and Clause Structure," 194.
[309] Porter, "Word Order and Clause Structure," 203.

Andrew Pitts's recent work in word order and clause structure bears the most resemblance to my own work.[310] After a survey of previous work in the area, some of which has been incorporated into the discussion above, Pitts turns to discussing the theoretical basis of his work.[311] The linguistic model informing Pitts's work is systemic-functional linguistics.[312] Pitts adopts it because it is the theory "forming the methodological basis for the OpenText.org database utilized" for Pitts's study of syntactical structure.[313] Adopting the model on which his database is based insulates Pitts from the charge that his analytical categories do not match his data and, furthermore, allow for reasoned comparisons between Pitts's work and other work using the OpenText database. This remedies one of the major shortcomings Pitts identified with previous work in this area: quantitative studies of word order have tended to work with smaller corpus sizes than would be ideal due to the difficulty of collecting the data, but simply aggregating the samples each researcher collected will not work because they have been collected on the basis of a variety of qualitative frameworks.[314] Pitts notes that word-order variation in languages like Greek is a matter of pragmatic function, indicating "prominence or focus (i.e. emphasis)."[315] He then makes reference to the theory of markedness developed by the Prague School of Linguistics. He particularly draws upon Batistella's division of linguistic markedness into three categories, semantic, cognitive, and distributional, aligning word order with the last of these.[316] Distributional

---

[310] Pitts, "Word Order," 311–46.
[311] Pitts, "Word Order," 313–21.
[312] Pitts, "Word Order," 313. Chapter 2 will cover this model in greater detail, since it is also the model I have chosen for my own work.
[313] Pitts, "Word Order," 313.
[314] Pitts, "Word Order," 312–13.
[315] Pitts, "Word Order," 314–15.
[316] Pitts, "Word Order," 315.

markedness relates to the frequency of a linguistic item: "the unmarked element or pattern will be more frequent than the marked element or pattern in a representative corpus of the language."[317] Pitts argues that determining marked word order "is not as easy as simply identifying the first structural element, however."[318] Marked patterns can occur at the level of the word group or the clause.[319] One also needs to consider whether or not the particular component is required or not.[320] Following Porter, Pitts applies Matthews' notion of codification to word order patterns: he treats patterns that appear more than 75 percent of the time as partially codified and 60 percent of the time as marginally codified.[321] The next theoretical consideration is the rank at which word order is investigated, either the word group or the clause.[322] The word group is what has traditionally been called the phrase, consisting of a head term and modifiers.[323] The patterns observed at clause level result from the ordering of four functional labels assigned to word groups: subject (S), predicator (P), complement (C), and adjunct (A).[324] In terms of corpora to interrogate, Pitts divides the NT into four categories: narrative (the Gospels and Acts), Pauline literature (including all thirteen letters specifically ascribed to Paul), "General Literature," the other letters (Hebrews–3 John), and lastly Revelation.[325] Pitts investigates both word group and clause patterns. He demonstrates several

---

[317] Pitts, "Word Order," 315.
[318] Pitts, "Word Order," 315–16.
[319] Pitts, "Word Order," 316; cf. Reed, "Theme," 87–89.
[320] Pitts, "Word Order," 316.
[321] Pitts, "Word Order," 316–17.
[322] Pitts, "Word Order," 317–21; cf. Dik, *Word Order*, 6.
[323] Pitts, "Word Order," 318–21.
[324] Pitts, "Word Order," 321.
[325] Pitts, "Word Order," 321.

significant patterns at the word-group level, noting that some of his results vary by register.[326]

Pitts's discussion of clause structure is more directly relevant to my project.[327] Pitts treats predicator and subject as "the basic elements of the Greek clause in the New Testament."[328] He restricts his analysis to "the relative positions of the stable constituents of the Greek clause to one another and to their complements," leaving aside adjuncts because they "mark further dependent structures" and instances of direct address because they function above the level of the clause.[329]

He suggests that clause structure patterns distribute differently among clause types.[330] As a result, he divides the report of his results into three sections, one for each type of clause.[331] I focus on his results for primary clauses here because that is the clause type on which my own project focuses. The search queries Pitts constructed for the OpenText.org database show a "clear tendency" for the predicator (i.e. verb) to precede its complement in primary clauses.[332] This is especially true for narrative literature like Mark and the documents to which I will compare it.[333] This indicates that the choice to place a complement before the predicator is significant.[334] Turning to the relative order of subjects and complements, Pitts finds that primary clauses tend to place the subject

[326] Pitts, "Word Order," 329. Register is a category in systemic-functional linguistics that describes the relationship between a text and the situation from which it sprang. Chapter 2 will discuss the relation of register to my project more specifically, but for the time being, the reader may consider it a rough synonym for genre.
[327] Pitts, "Word Order," 330–39.
[328] Pitts, "Word Order," 330.
[329] Pitts, "Word Order," 330.
[330] Pitts, "Word Order," 330.
[331] Section 2.2.1 covers primary clauses (Pitts, "Word Order," 331–34), section 2.2.2 covers secondary clauses (Pitts, "Word Order," 335–37), and section 2.2.3 covers embedded clauses (Pitts, "Word Order," 337–39). Chapter 3 will cover these clause types in more detail.
[332] Pitts, "Word Order," 333.
[333] Pitts, "Word Order," 333.
[334] Pitts, "Word Order," 333.

before the complement.[335] Finally, primary clauses exhibit a "slight tendency" to place the subject before the verb.[336]

Pitts draws out several implications of his work for further research in this area. First, with regard to grammar and syntax proper, he points out that Greek word order is not as random as some previous researchers have suggested.[337] The constraints, however, do not operate exactly the same for every sort of literature.[338] This implies that "the degree to which a structure is marked will be largely constrained by the individual discourse or (perhaps) corpus.[339] Second, Pitts points out that, insofar as "discourse analysis analyzes linear structures in terms of information flow using the terminology of theme/rheme," his results "push this analysis forward through gaining a clearer understanding of how syntactic clause constituents work with the functional pragmatic categories of discourse analysis."[340] He suggests the basic patterns are placing the verb first as an unmarked order and placing the subject first as a marked order.[341] The presence of a fronted complement "is especially marked in narrative, placing prominence upon the semantic content of the complement relative to the predicate."[342]

Building on the work of Porter and Pitts, Chris S. Stevens recently published an analysis of Philippians in which clause structure is one of three "objective grammatical means" to which the article's subtitle refers.[343] According to Stevens, clause structure analysis based on SFL offers two "distinct advantages": (1) it "is concerned with purely

[335] Pitts, "Word Order," 334.
[336] Pitts, "Word Order," 334.
[337] Pitts, "Word Order," 340.
[338] Pitts, "Word Order," 341–42.
[339] Pitts, "Word Order," 342.
[340] Pitts, "Word Order," 344.
[341] Pitts, "Word Order," 344.
[342] Pitts, "Word Order," 345.
[343] Stevens, "Objective Grammatical Means," 327–49.

quantifiable grammatical features" and (2) it provides "foundational statistics" for identifying "noteworthy anomalies or clusterings of features."[344] Pursuing the second of these advantages involves examining a whole discourse and then comparing the results to those obtained for the portion of that discourse with which one is concerned.[345] For example, Stevens shows that the frequency with which explicit subjects and verbs occur in Philippians 2 is "grammatically consistent" with the letter as a whole; he infers from this that "the author has not chosen to highlight a feature in Phil 2 simply by a drastic alteration in the presence or absence of a particular constituent."[346] By contrast, however, he is able to show that Philippians 2 does show significantly different patterns in the positioning of these elements within clauses. Subjects are placed at the front of the clause roughly half as often in Philippians 2 than they are in the letter as a whole.[347] Similarly, verbs occur at the end of a primary clause roughly half as often in Philippians 2 compared with the letter as a whole.[348] As an example of moving beyond statistics to assessing the possible motivations that produced the divergences, Stevens points to Philippians 2:4–6 where there are three clauses in a row that begin with explicit subjects, a sequence occurring nowhere else in the letter, as a "highlighted peak."[349]

Conclusion

This literature review surveyed several areas relating to my attempt to ascertain which component of the comparative corpus I have assembled bears the most resemblance to the Gospel of Mark. The notion of using a comparative corpus to adjudicate the genre to

[344] Stevens, "Objective Grammatical Means," 336.
[345] Stevens, "Objective Grammatical Means," 338.
[346] Stevens, "Objective Grammatical Means," 338.
[347] Stevens, "Objective Grammatical Means," 339.
[348] Stevens, "Objective Grammatical Means," 339–40.
[349] Stevens, "Objective Grammatical Means," 341.

which a given document belongs only makes sense given a particular understanding of the concept of genre, so the first step was to defend my adoption of that understanding. Having determined that comparison with contemporaneous documents is a valid means of ascertaining the genre to which a document belongs, the question naturally arises what sort of analogies are available in the first-century milieu. As a result, I devoted some space to describing history, biography, and the ancient novel, which not coincidentally serve as the basis of my comparative corpus.

Following this, I described three stages in the progress of scholarly investigation into the question of the gospel genre. The first stage, during which biography was the unquestioned analogy for the gospels, ran from the second century through the beginning of the twentieth century. The dominant perspective within the second stage, which consisted of roughly the first two-thirds of the twentieth century, was that no appropriate analogues for the Gospels existed in the first-century literature.

Within the third stage, I singled out two figures for special consideration, Richard Burridge and David Mealand, because their quantitative approaches resemble my own process more closely than other contributors to the debate. Building on work by James Libby and Andrew Pitts, I noted some methodological problems with the work of both these scholars. Preeminent among these is failing to explain how the interesting distributional differences they observed contribute to the question of the gospel genre.

To date, investigations into the nature of the gospel genre seem to have foundered on one side or the other of the following divide: “On the one hand, a qualitative generalization without a quantifiable prediction is hard to falsify or replicate, but on the other, a quantitative generalization without a qualitative motivation or implication is hard to interpret."[350] Modern linguistics, especially the form known as corpus linguistics, provide falsifiable, replicable quantitative generalizations.[351] As such, it "affords us with a higher degree of comparability, objectivity, and replicability" than traditional, qualitative methods do.[352] Likewise, modern linguistic models offer qualitative accounts of the reasoning and implications of these quantitative generalizations. To these twin topics of corpus linguistics and linguistic models I now turn.

[350] Cantos, "Use of Linguistic Corpora," 100; cf. Jenset and McGillivray, *Quantitative Historical Linguistics*, 3.

[351] According to Terttu Nevalainen and Helena Raumoulin-Brunberg ("Historical Sociolinguistics," 23), genre is "one of the most researched topics in English historical sociolinguistics."

[352] Gries, "Elementary Statistical Testing," 361.

CHAPTER 2: MAPPING TEXTS (METHODOLOGY)

The literature review in Chapter 1 surveyed both intuitive and quantitative attempts to place the canonical gospels within the literary milieu of the first-century Greco-Roman world. It described the logjam of competing claims that traditional, intuitive methods have produced. I suggested that quantitative methods have the potential to break the logjam and forge a greater degree of consensus by providing hard data for evaluating the traditional hypotheses. To date, however, quantitative methods used by Richard Burridge and David Mealand have failed to realize this potential. In my view, at least one cause of this failure is that neither of them cogently explains how the move from concrete features of a text to statements about its genre, which is an abstract property. I aim to push the discussion forward by providing such a link.

I find the metaphor of maps helpful as a framework for describing the process of analyzing texts.[1] Several components must come together to produce a useful map. First, useful maps are consistently drawn according to a particular projection. Second, useful maps result from a careful surveying process that quantifies the features the map is to represent. Third, useful maps represent the lay of the land in a particular, defined area. I now turn to describing in further detail each of these components and how the relate to textual analysis.

A map projection is a tool for compressing the spherical Earth into two dimensions. As one might expect from a process that compresses three dimensions into two, every map projection distorts reality to some degree. Different projections

[1] I first encountered this metaphor in the work of Matthiessen (*Lexicogrammatical Cartography*, passim), and Webster ("Introduction," 2–3) mentions that M. A. K. Halliday, about whom I will have much more to say below, has used it as well.

accomplish their tasks in different ways, resulting in different sorts of distortion. The bewildering array of possible projections means one can typically employ one whose particular distortion does not materially affect the task at hand. Likewise, language is a complex phenomenon that defies comprehensive description.[2] Some form of abstraction and idealization is necessary to render its complexity manageable.[3] The tools of this abstraction, analogous to map projections, are called linguistic models; they are indispensable to the process of textual analysis.[4]

Granted, the majority of interpreters do not choose their linguistic model consciously, and thus depend on an ill-defined, intuitive model, but they are using a model nevertheless.[5] Human communication occurred long before modern linguistics was around to describe it systematically, which testifies to the effectiveness of humanity's intuitive, *ad hoc* linguistic models under at least many circumstances, but explicit theoretical reflection on language offers several advantages, particularly in bringing differing preconceptions to the surface.[6] Intuitive models work in the vast majority of cases where those communicating are on the same page, so to speak, but communication can quickly break down when they are not. Perhaps an analogy with maps will help. A map intended to direct a friend to a particular place in town likely does not need more than a few major streets and landmarks indicated because the intended recipient shares

---

[2] Cf. Gries, "Sources of Variability," 5.

[3] Cantos, "Use of Linguistic Corpora," 100; Halliday, *Language as Social Semiotic*, 52; Halliday, "Theory to Work," 131–32.

[4] Cf. Halliday, "Machine Translation," 23; Jenset and McGillivray (*Quantitative Historical Linguistics*, 4–6) divide linguistic models into several categories, including some offering "statistical analysis of annotated corpora that are enriched with part-of-speech information or syntactic annotation, in order to draw conclusions about usage, grammar, or language change" (6). This aptly summarizes my project, which attempt to analyze a syntactically annotated corpus to draw a conclusion about linguistic usage in particular situations, i.e. register variation.

[5] Halliday, "Language in Relation to Fuzzy Logic," 197.

[6] Cf. Halliday, *Grammar*, xxvix; Stevens, "Clause Structuring," 65.

the necessary background, whereas the map on which an orienteer relies for traversing country he or she has not visited previously needs far more detail. Likewise, I would argue that at least part of the reason that different interpreters come to differing understandings of the genre of the canonical Gospels is that each interpreter has differing pictures of the layout of town, so to speak, so the orienteer's map is the better model for the situation.

As with map projections, the landscape of modern linguistics offers a variety of competing models I could potentially use to guide my study.[7] This embarrassment of riches has allowed me to choose a model I think is particularly suited to answering my question. I considered three models, eventually opting for systemic-functional linguistics (SFL). The first section below describes each option and why I think SFL is optimal for my particular project. The second section below goes into more detail about SFL.

Moving to the second analogy with physical maps, useful textual maps also result from a consistent surveying process. The third section below covers the surveying process I used for this study.

Lastly, useful textual maps cover a particular area. As with normal maps, the size of this area is inversely proportional to the level of detail. If a map of the entire world is to remain a reasonable size, it cannot show the locations of individual houses. On the other hand, a map of a town could potentially show their locations but would say nothing about the countless other towns on Earth. Thus, the appropriate map for any particular task depends on whether the task calls for observing lots of things or fewer things in more detail. A task will often call for a mixture of the two (several things at a reasonable level

[7] Cf. Sampson, *Schools*, passim.

of detail). The textual maps constituting my project are such a case: they need to be detailed enough to be convincing, but general enough that I can generate a sufficient number of them to represent a large swath of the first-century literary milieu. The fourth section below describes the principles underlying this process.

Potential Linguistic Models for Studying the Gospel Genre

This section describes the linguistic models I considered and the reasons I opted for SFL. The first subsection lays out the criteria that informed my choice of model. The next three subsections discuss the three candidate models for this project: Transformational Grammar, Construction Grammar, and SFL. Each model description lays out the theory and a report of its application to the Greek of the New Testament.

Criteria for Choosing a Linguistic Model

Both practical and theoretical criteria impinge on my choice. In terms of practical features that will facilitate will facilitate my work, the ideal model would have a pedigree of application to the New Testament. Secondly, it would also offer a set of easily replicable quantitative categories so that comparisons between texts analyzed with these categories will be reliable and easy to implement with statistical tools. The central consideration, however, is theoretical in nature, the very one I identified in Chapter 1 as the failure point for previous quantitative work on the gospel genre: how does one convincingly move from a text’s observable features to its abstract, immaterial properties, such as genre? I now turn to discussing this vital question; this theoretical discussion and the practical considerations just mentioned provide the framework for the evaluation of candidate models in the remainder of the section.

The central difficulty in convincingly relating observable features of a text to statements regarding its genre derives from the fact that these entities are phenomena of two distinct kinds.[8] The distribution of forms within a text is a linguistic phenomenon. Genre, on the other hand, is a cultural phenomenon.[9] Indeed, one recent conspectus of modern genre theory claims that genre is in danger of becoming "a cultural buzzword."[10] Language clearly plays a pivotal role in the development and transmission of culture, so the two are at least somewhat related, but any attempt to account thoroughly for a cultural phenomenon by means of language will clearly have to take the social context in which language is used because language itself does not exhaust culture.[11] That is to say, there are a number of aspects of "what language means" that in fact require one to "move outside language in order to explain" them.[12] In fact, the elements of culture that authors take for granted, rather than explicitly state, are precisely those most directly related to my project: a fairy tale does not need to begin "this is a fairy tale"; beginning with the words "once upon a time" will do the trick of orienting the reader for what is to come.

Moving beyond language itself entails taking a position on how linguistics relates to other disciplines. Models differ greatly with regard to the way in which and the degree to which they account for the link between language and the external world. At the risk of oversimplification, particular linguistic models tend to position themselves relative to one

---

[8] For some linguists, this means differentiating between genre and "text type": "Genre categories are determined on the basis of external criteria . . . whereas text types refer to classes of texts that are grouped on the basis of similarities in linguistic form, irrespective of their genre classifications" (Cantos, "Use of Linguistic Corpora," 116).

[9] Adams provides a helpful survey of ancient and modern perspectives on genre (*Genre*, 26–67).

[10] Duff, "Introduction," 2.

[11] Hasan, "Language and Society," 6; Hasan, "What's Going On," 83. Coming at the issue from literary criticism, Duff ("Introduction," 2) argues that the concept of genre raises larger question of the "organization and transmission of knowledge and the dynamics of cultural change."

[12] Halliday and Hasan, *Language*, 4.

of two other disciplines: as Halliday puts it, "you either say with Chomsky that linguistics is a branch of theoretical psychology, or—which is equally valid—that linguistics is a branch of theoretical sociology."[13] That is to say, models either focus on the mental processes within individual language users, or they focus on how those language users interact with each other, at least partially enacting and maintaining a social structure via language.[14]

This dichotomy between psychologically-oriented linguistic models and sociologically-oriented ones is a fundamental, irreconcilable opposition. One could theoretically find common ground on the level of description, but explanation of phenomena requires a decisive step in one direction or the other, because psychologically-oriented and sociologically-oriented explain very different things and, therefore, explain them in very different ways. No model can pay equal attention to every question. After all, one of the main purposes of a model in the first place is to limit the data with which a researcher concerns himself or herself to that which assists with a particular question. The important factor in choosing a model, then, is to make sure that one's choice of model matches the question in which one is interested. To use an everyday analogy, I might eventually be able to turn a screw with the claw-end of a hammer, but—to say the very least—a hammer is not the most efficient tool for the job.

Thus, the question of what model to use for this project boils down to: is genre an individualistic phenomenon or a social one? The latter is clearly the case. The notion of genre is a classic case of exploiting social conventions. Shared cultural knowledge of the

[13] Halliday, *Language as Social Semiotic*, 57; Halliday, *Grammar*, xxviii.
[14] Halliday and Hasan, *Language*, 4; cf. Halliday, *Language as Social Semiotic*, 4. Porter and Pitts term these two groups as "generative" and "functional," respectively ("NT Greek Language," 230).

expectations associated with a genre and the signals that evoke them allows the author to enter into an implicit contract with the reader: if he or she provides the signals, the reader can expect them to fulfill the concomitant expectations.[15] Or, viewed from the other side, evoking genre expectations allows an author to inform the reader of the "rules of the code" governing "how the author asks the reader to approach" the text.[16] The rules are implemented through language, and thus language bears witness to them, but the study of language alone cannot describe them adequately, for they themselves are extralinguistic. Therefore, a useful linguistic model for my study will be one that studies language along with its relation to the external world.[17]

Interpreters of the New Testament have applied several linguistic models to varying degrees of success.[18] In particular, three models deserve mention here because of their standing in linguistic investigations of the New Testament: transformational grammar, case grammar, and systemic functional linguistics. The following three subsections briefly summarize this history before asking three questions derived from the practical and theoretical considerations of model selection mentioned above. These questions, listed in order of importance, are: (1) Does the model in question conceive of language psychologically or sociologically? (2) Does it provide clear, replicable categories? (3) Does this model produce data in a quantitative form amenable to statistical analysis?

---

[15] Cf. Porter, "Multidisciplinary," 106–07
[16] Adams, *Genre*, 1; cf. Burridge, *Gospels*, 25–31.
[17] Halliday, *Language as Social Semiotic*, 187; cf. Chambers, "Language Variation," 12.
[18] Porter and Pitts, "NT Greek Language," 214–55.

Transformational Grammar

The work of Noam Chomsky so dominated linguistics in the twentieth century, particularly in North America, that for many scholars in both linguistics and neighboring disciplines "linguistics is largely synonymous with the Chomskyan revolution."[19] One testament to Chomsky's towering "standing in the discipline" is that Sampson takes for granted that his readers will already be familiar with Chomsky's work, even though Sampson's discussion of Chomsky occurs several chapters later.[20] Clearly, such a pedigree entitles transformational-generative grammar to at least be considered as the model I will use to structure my quantitative investigation of genre.[21] I now turn to discussing the essence of the theory before describing its application and posing the three questions for model evaluation developed in the last subsection.

Understanding Chomsky's revolutionary views and impact is difficult without understanding the scholarly milieu out of which he arose. This would likely be true in any case, but it is particularly important in this particular case because Chomsky is consciously breaking the paradigm of previous linguistic study by importing insights from the two other major scholarly disciplines with which he was associated, namely mathematics and philosophy.[22] Thus, an appreciation of Chomsky's contribution rests on laying out both his reason for blazing a new trail and the new tools with which he blazed it.

---

[19] Hasan, "Meaning," 38; cf. Halliday, *Grammar*, xxviii–xxvix. Sampson shows the aptness of the terminology of "revolution" by noting that like "books published in the Soviet Union on the most abstract academic topics once had to begin with a ritual obeisance to the guiding genius of Stalin, so nowadays even scholars researching aspects of language which have very little connection with Chomsky's work often feel obliged to claim publicly that their writings exemplify the Chomskyan paradigm" (*Schools*, 130).

[20] Sampson, *Schools*, 247n12.

[21] Cf. Thomas, *Key Thinkers*, 249–59.

[22] In particular, Sampson (*Schools*, 13) notes that Chomsky acknowledges a debt to Wilhelm von Humboldt and the French rationalists of the seventeenth century.

Chomsky emerged onto a scholarly landscape dominated by the American descriptivist school, which—as its name suggests—primarily concerned itself with describing languages.[23] While the model ultimately derived from the work of Franz Boas, a German-American anthropologist who conducted a survey of the various languages spoken by tribes of Native Americans living in the United States and Canada and influenced entire generations of American linguists, the work of two other scholars, Leonard Bloomfield and Zellig Harris, better explains that against which Chomsky's work reacts.

Bloomfield's claim to fame is managing "to promote and codify the Descriptivist tradition" and "to organize linguistics as a profession."[24] The former of these two contributions makes Bloomfield's work a convenient reference point for the thought of the movement, but I am more particularly concerned with the latter contribution here. He devoted a great deal of attention to ensuring that the scholarly community recognized linguistics as a science. Within his context, in which logical positivism was the reigning paradigm among philosophers of science, gaining recognition as a science involved making one's study reducible to logic, sense data, or some combination of the two. Fortunately for Bloomfield and his quest—although, in my view, perhaps unfortunately from the perspective of progress within the discipline—psychologists faced with this same scholarly pressure had already met the challenge by creating the model of behaviorism.[25] This ready-made framework allowed Bloomfield to classify linguistics as

---

[23] The following description builds on Sampson's discussion (*Schools*, 57–80).

[24] Sampson, *Schools*, 62; cf. Thomas, *Key Thinkers*, 167–73.

[25] In the course of Sampson's later discussion of the Prague School, a group of European linguists roughly contemporary with Bloomfield, who have been significant for linguistic study of the New Testament as well (cf. Nylund, "Prague School," 208–21), he contrasts their attitude towards this question with Bloomfield's: the linguists of the Prague School were perfectly happy to remain on the arts side of the

"a branch of psychology," specifically behavioristic psychology.[26] Adopting a behavioristic perspective had at least two effects on the descriptivist movement, one positive and one negative. On the positive side, so long as the analysis remained at levels where there was a clear connection between stimulus (input) and response (output), e.g. phonology and morphology, focusing one's description of a language on actual use of that language, i.e. on observable behavior, rather than the language user's subjective intuitions about his or her language, had a positive impact in terms of methodological rigor, forcing descriptions to become "genuinely scientific rather than a bastard mixture of statements testable against observation versus statements that had to be taken on faith."[27]

However, at times—particularly when it tried to tackle semantics—linguistic behaviorism went to seed: some behaviorists "confused the methodological issue," i.e. the positive aspect above, "with a matter of substantive belief," leading them to write "as if belief in the existence of minds and mental activity were on a par with belief in the existence of a water-god who is angry when the sea is rough."[28] Two particular instances of linguistic behaviorism running amuck, which are particularly relevant to understanding the background of Chomsky's thought, are: (1) the atomistic focus on the structure of particular languages to the exclusion of theoretical reflection on the nature of language as a whole or the communicative process, and (2) the belief that linguists could develop procedures "to derive the correct grammar of a language from a corpus of observed data

arts/science divide and, thus, felt no pull as a result of logical positivism (*Schools*, 112). How different might the history of linguistics be if Bloomfield and the Descriptivists had taken the same tack?

[26] Sampson, *Schools*, 64.

[27] Sampson, *Schools*, 69.

[28] Sampson, *Schools*, 66.

in a purely mechanical way."[29] Being a behaviorist, positivistic theory, Bloomfieldian descriptivism was unwilling to posit linguistic universals because they could not be observed directly, short of a complete enumeration of all human language: "Features which we think ought to be universal may be absent from the very next language that becomes accessible."[30] If there are no universals, there is no need to theorize, since the purpose of a scientific theory is to explain similarities between phenomena., so the descriptivists contented themselves with making observations about the structure of various languages but felt no need to generalize these into a coherent, integrated theory that explained language as a whole.[31]

The second consequence of rampant behaviorism, the idea that linguists ought to be able to develop procedures to analyze grammar mechanistically, is where we encounter the latter of the above-named descriptivist scholars, Zellig S. Harris.[32] Harris' relevance to Chomsky's thought is rather more direct and personal than Bloomfield's; Harris introduced Chomsky to linguistics while the latter was studying at the University of Pennsylvania where Harris taught.[33] Harris was one of the few descriptivists to tackle syntax, and he did so by dividing morphemes into groups based on their distribution in text relative to other morphemes.[34] Having comprehensively grouped the morphemes into distributional classes, Harris then came up with rules that describe how these

---

[29] Sampson, *Schools*, 75.

[30] Bloomfield, *Language*, 20.

[31] Fillmore ("Case," 1) humorously comments: "The writer recalls a Linguistic Institute lecture of not many summers ago in which it was announced that the only really secure generalization on language that linguists are prepared to make is that 'some members of some human communities have been observed to interact by means of vocal noises'."

[32] Cf. Thomas, *Key Thinkers*, 171.

[33] Sampson, *Schools*, 130.

[34] Lyons (*Introduction*, 183–84) discusses the concept of "morpheme," and its relation to "words."

distributional classes interact, and the complete set of these rules constituted his account of syntax.[35]

Moving beyond the preceding descriptivist agenda of simply labeling observed phenomena, Chomsky aimed to account for the effectively universal ability of humans to communicate via language, what he termed "competence."[36] Conceiving of competence as an innate attribute of humans as a species, Chomsky infers that grammar, as the description of competence, must also be universal. This inference leads to a further one: the irrefutable proof of differences in surface structure entails that competence, which is by definition universal, describes something other than the observable surface structure of the text.

Despite the dominance of Chomsky's work in mainstream linguistics, few have attempted to apply it to the Greek of the New Testament.[37] Early on, a German scholar attempted to illustrate the usefulness of generative syntax in the interpretation of the New Testament from a discussion of two key Pauline texts.[38] Two major works followed in English. The first of these, appearing in 1981, Daryl D. Schmidt's *Hellenistic Greek Grammar and Noam Chomsky: Nominalizing Transformations* attempts to develop rules that explain how a clause can fill slots in another, larger clause typically filled by a noun; such an embedded clause has been "nominalized."[39] Later, Michael Palmer applied a form of generative grammar known as X-bar syntax to several sections of Luke–Acts and the Pauline letters.[40] Most recently, Robert Crellin has melded two methods, one of which

---

[35] Sampson (*Schools*, 134–37) gives a fuller account of this process.
[36] Cf. Hasan, "Language and Society," 24.
[37] For what follows, cf. Porter and Pitts, "NT Greek Language," 231–33.
[38] Wonneberger, *Syntax*, passim.
[39] Schmidt, *Hellenistic Greek Grammar*, passim.
[40] Palmer, *Levels*, passim.

is reminiscent of construction grammar and the other is one of the more recent forms of transformational grammar, Government-Binding theory.[41]

The failure of Chomsky's work to catch on in New Testament studies likely stems from two factors. First of all, as several scholars have noted, until recently very few interpreters of the New Testament interacted with modern linguistics much at all and—even now—the uptake of such methods is perhaps not what it should be.[42] That, in and of itself, says more about the shortcomings of investigation into the Greek of the New Testament than it does about the shortcomings of transformational-generative grammar. However, far more problematic is the poor record transformational-generative grammar has at describing what K. Hale describes as "non-configurational languages," i.e. languages that have relatively few constraints on word order, allow for the possibility of elements embedded within the surface structure of other elements, and whose verbs do not require an explicit subject.[43] Hale was working on a language called Warlpiri, but others have followed up and validated his findings for a number of other languages spoken across the globe. The gist of their work is that the three characteristics of non-configurational languages violate the preconditions of traditional Chomskyan syntax, raising substantial questions about just how universal Chomsky's universal grammar is.[44] Granted, more recent approaches have attempted to deal with these problems, but those applying generative syntax to the New Testament have used the older models that Hale and others have discredited, even after alternatives were available.[45]

---

[41] Crellin, *Syntax and Semantics*, 72. In terms of the relation between Crellin's method and construction grammar, compare Crellin's table of "semantic roles" and the case inventories of case grammarians (74).

[42] E.g. Palmer, "How do We Know," 154 (esp. n 1);

[43] Hale, "Warlpiri," 5–47.

[44] Cf. Porter and Pitts, "NT Greek Language," 232.

[45] Porter and Pitts, "NT Greek Language," 233.

Transformational grammar falls short as a model for this project.[46] It is the epitome of a psychologically-oriented model, which makes convincingly accounting for genre variation difficult. Likewise, it prioritizes introspective discovery procedures that do not lend themselves to clear, replicable categories or statistical analysis. Lastly, its application to Hellenistic Greek is problematic at best.

Case and Construction Grammar[47]

Case grammar developed from the work of Charles J. Fillmore.[48] This is another generative theory, developed in reaction to Chomsky's.[49] It also borrows some ideas from the Prague School. Danove characterizes construction grammar as "a descriptive, non-transformational grammar which renders the locutions of a language systematically according to their syntactic and semantic properties."[50] Construction Grammar is descriptive in the sense that it "proceeds in immediate reference to received locutions," i.e. actual instances of text.[51] It is non-transformational because there are "no intermediate stages in the generation of a sentence beyond those that lead directly to its surface structure."[52] It is a grammar "in that it constitutes a formalized system for characterizing the linguistic processes and resources of language."[53] This grammar seeks to "identify a particular set of sentence elements and detail the syntactic and semantic constraints on these elements."[54]

---

[46] Cf. Chambers, "Language Variation," 7–12.
[47] Porter and Pitts, "NT Greek Language," 224–30.
[48] Fillmore, "Case," 1–88.
[49] Danove, *Linguistics and Exegesis*, 14; Porter and Pitts, "NT Greek Language," 224–25.
[50] Danove, "Theory of Construction Grammar," 120.
[51] Danove, "Theory of Construction Grammar," 120.
[52] Danove, *Linguistics and Exegesis*, 15.
[53] Danove, *Linguistics and Exegesis*, 13.
[54] Danove, *Linguistics and Exegesis*, 16.

However, with regard to its relative reception within mainstream linguistics and New Testament studies, we find the reverse of what was the case for transformational-generative grammar: whereas the latter dominated—at least for a period—mainstream linguistics but largely failed to catch on in New Testament studies, the former continues to exercise influence in investigation of the Greek New Testament long after mainstream linguistics has largely discarded it. The two major proponents of case grammar in New Testament studies are Paul Danove and Simon Wong.[55]

As Danove implements it, construction grammar divides the elements of a text into three categories: (1) "predicators," i.e. words that "either require or permit the presence of other phrasal elements," (2) "arguments," i.e. words that are required by a predicator, and (3) "adjuncts" i.e. words that are permitted, but not required, by a predicator.[56] Each predicator has certain arguments associated with it according to the meaning of the predicator in question.[57] For instance, Danove indicates that the verb δίδωμι has three arguments, "for a minimal description of this act must reference three participants," namely the giver, the recipient, and the thing being given.[58] Adjuncts, on the other hand, simply "add to the meaning of the phrase but are not required."[59] Danove has applied this model to a variety of texts, especially the Gospel of Mark, and also a variety of broader grammatical questions.[60]

---

[55] Porter and Pitts, "NT Greek Language," 228–29.
[56] Danove, *Linguistics and Exegesis*, 17.
[57] Danove, *Linguistics and Exegesis*, 17–18.
[58] Danove, *Linguistics and Exegesis*, 18.
[59] Danove, *Linguistics and Exegesis*, 20.
[60] E.g. Danove, *End*, passim; Danove, *Linguistics and Exegesis*, passim; Danove, *Rhetoric and Characterization*, passim; Danove, "Theory of Construction Grammar," passim; Danove, "Verbs of Experience," passim.

SFL

SFL arose out of the work of the British linguist J. R. Firth and especially M. A. K. Halliday, one of his students.[61] In contrast to the models discussed previously, transformational-generative grammar and case grammar, SFL is not a generative theory. Rather, its name specifies what sets it apart from other linguistic models: it is both systemic and functional.[62] SFL is systemic because it approaches language as a semiotic system, i.e. a system capable of "carrying or creating meaning."[63] It is functional because it conceives of language as a "symbolic resource," whose form derives from the job it is called upon to accomplish, namely transferring meaning between two or more people within a social context who are trying to communicate with each other.[64]

Semiotics, the "investigation of how things mean," is not a new idea in language study.[65] In fact, the term itself derives from Greek equivalents that the Stoic grammarians introduced as far back as the second century BCE to refer to the distinction between a word as a sign and the meaning associated with it.[66] However, the Stoics viewed the connection between a sign and its meaning as being one that existed in and of itself apart from a sign's relationship to other signs.[67] Ferdinand de Saussure, who pioneered the discipline of modern linguistics, discussed how signs related to each other in great detail, but he did not allow this "very strong conception of language as a system of

---

[61] Cf. Thomas, *Key Thinkers*, 238–43.

[62] Matthiessen et al., *Key Terms*, 216; cf. Halliday, "Pinpointing," 150; Hasan, "Language and Society," 29.

[63] Matthiessen et al., *Key Terms*, 194; cf. Halliday, "Working," 38; Halliday, "Why," 72.

[64] Halliday, *Language as Social Semiotic*, 4–5; cf. Halliday, *Grammar*, xxvii–xxvix; Halliday and Hasan, *Language*, 44.

[65] Cf. Halliday, "Towards Probabilistic," 43.

[66] Halliday and Hasan, *Language*, 3.

[67] Halliday and Hasan, *Language*, 3.

relationships" to displace what Halliday characterizes as a "rather atomistic conception of the linguistic sign."[68]

Halliday's contribution to the debate, then, is to take the leap to which—Halliday believes—the logic of de Saussure's position would have led him, if he had followed it consistently: Halliday broadens semiotics to include "the study of sign systems—the study of **meaning** in its most general sense;" in this view, individual signs function as the "external form of output" for systems of meaning.[69] For example, language is a semiotic system whose distinctive output is "a system of ***wordings*** (words together with associated structural patterns)."[70]

Moreover, beyond simply calling for scholars to broaden their idea of "sign," Halliday also makes the further move—even more significant for my purposes—of tying systems of meaning together and thus viewing culture as a "set of semiotic systems."[71] Although language is only one member of the large set of meaning-bearing systems, it is one of particular importance for its enabling of others: many semiotic elements of human culture, such as art or music, depend on human language for their transmission.[72]

Thus, Halliday classifies language as a "social semiotic;" this term entails that language carries meaning by virtue of its "sociocultural context, in which the culture itself is interpreted in semiotic terms."[73] Treating both language and its environment in semiotic terms allows them to shed light on each other.[74]

---

[68] Halliday and Hasan, *Language*, 3.
[69] Halliday and Hasan, *Language*, 4; cf. Halliday, "Language in Relation to Fuzzy Logic," 198-99.
[70] Halliday, "Language in Relation to Fuzzy Logic," 199.
[71] Halliday and Hasan, *Language*, 4; cf. Hasan, "What's Going On," 80.
[72] Halliday, "Language in Relation to Fuzzy Logic," 200.
[73] Halliday, *Language as Social Semiotic*, 2; cf. Halliday and Hasan, *Language*, 4.
[74] Halliday and Hasan, *Language*, 11–12.

More specifically, Halliday refers to his approach as a sociological approach, rather than just a social approach, because he attempts to tie his account of language to "social contexts that are themselves of significance" with criteria "based on some theory of social structure and social change."[75] Preeminent among the social factors that undergird SFL's approach is the fact that communication clearly occurs: if communication occurs, then there must be some system of norms that allows for the transfer of meanings from one person to another. According to SFL, this system of norms results from "interrelations among the three levels of" social processes (whose linguistic embodiment is the text), the situation in which the interaction takes place, and the linguistic system itself.[76]

Considering language as a resource for exchanging meanings in a social context places the brunt of the responsibility on how the forms that make up a language relate to each other; to use linguistic terms, a social approach is paradigmatic, rather than syntagmatic, in orientation.[77] As Halliday puts it, "Text is meaning, and meaning is choice, a current of selections each in its paradigmatic environment of what **might have been** meant (but was not)."[78] In other words, for SFL, syntagmatic order is relevant because it sequences paradigmatic choices that carry meaning.[79]

Of the options I considered, SFL is clearly the "model of choice" for my inquiry.[80] Attention to both the "internal organization" and the "external relations" of language has from the beginning characterized the "unique programme of research"

[75] Halliday, *Language as Social Semiotic*, 35.
[76] Halliday, *Language as Social Semiotic*, 64.
[77] Halliday, *Grammar*, xiv; Halliday, *Language as Social Semiotic*, 40–41; Matthiessen et al., *Key Terms*, 214.
[78] Halliday, "Text as Semantic Choice,"48; cf. Halliday, "Information and Meaning," 57.
[79] Halliday, "Theory to Work," 133.
[80] I am borrowing this turn of phrase from Webster ("Introduction," 1).

undertaken by practitioners of SFL.[81] Thus, using SFL remedies the primary failing I identified above with previous attempts to apply linguistic data to understanding the genre of the Gospels, i.e. their failure to explain how their observations count as evidence for the genre question.

Detailed Look at SFL

Different practitioners of SFL view genre differently.[82] The option I find most persuasive builds upon the recognition that genre operates as a cultural system that carries meaning, i.e. a semiotic system, that one can study by means of the same theoretical tools for studying other semiotic systems.[83] SFL theory offers three perspectives from which to study semiotic systems: (1) considering how the output of the system relates to the social environment from which it sprang (using a theoretical tool called “stratum”), (2) considering how one can build up a picture of the total system from individual outputs (using a tool called “instantiation phase”), and (3) studying how language use accomplishes the general tasks (called “metafunctions”) that all languages must somehow accomplish in order for communication to occur.[84]

Genre relates directly to the social environment of language use (one is more likely to encounter a business letter in certain social environments and a fairy tale in others), so the perspective of stratum seems promising for understanding genre. Likewise, building up a picture of the system through exposure to particular instances is precisely

[81] Hasan, “Language and Society,” 5–6.

[82] Halliday (e.g. *Language as Social Semiotic*, 145) represents genre as an aspect of register variation, on which see below, whereas Martin (e.g. *English Text*, 500–07) prefers to make genre a separate plane of which register variation is the expression.

[83] Matthiessen et al., *Key Terms*, 106–07.

[84] Matthiessen et al., *Key Terms*, 38.

how readers acquire genre expectations, so instantiation phase seems likely to be profitable as well.

As the above discussion implies, many questions—genre being among them—require considering how the three perspectives interact.[85] In fact, the primary use for metafunction in my project is how it ties the other two perspectives together.[86] Also, more prominently, Halliday models his equivalent for genre, "situation type," in terms of the intersection of instantiation and stratification.[87] Before further discussing these intersections, however, I turn to describing each dimension on its own terms.

Instantiation

The evolution of language comes about because it continually interacts with its environment, namely culture.[88] Halliday maintains that instances of communication are the means by which language evolves.[89] This follows directly from appraising language as a system: each instance is part of the "universe of countless" instances that make up "the system of language."[90] In other words, instantiation phase is the use of inductive reasoning to infer the meaning potential of language from the manifestations of that potential in actual instances of communication.

Halliday's notion of instantiation phase arose because he sees an underlying system behind any use of language: "each occurrence of a sign is an instance of (a term in) an underlying system."[91] Each individual instance contributes to—and, thus,

---

[85] Matthiessen et al., *Key Terms*, 39.
[86] Cf. Matthiessen et al. *Key Terms*, 104.
[87] Halliday, "Towards Probabilistic," 60; cf. Matthiessen et al., *Key Terms*, 125.
[88] Halliday, "Towards Probabilistic," 44.
[89] Halliday, "Towards Probabilistic," 44.
[90] Halliday, "Towards Probabilistic," 44.
[91] Halliday, "Driving Force," 167.

potentially disturbs—the system, although the disturbance will typically be small and cancelled out by disturbances from other instances.[92] Sometimes, however, disturbances in the same direction accumulate to the point where one can recognize evolution within the linguistic system; this is one potential avenue of studying language from the perspective of instantiation.[93]

More relevantly for my study, however, the notion of instantiation offers a "window on the system."[94] Each instance shows a different (though potentially overlapping) portion of the system, and thus a reliable picture of the system should emerge from a large collection of instances.[95] This collection of instances, called a corpus, can "enable us to see more closely, and more accurately, into that underlying system."[96] The process of assembling a corpus mirrors the process by which a child learns to use his or her native language: in both cases, knowledge of the system comes about as a result of instances of communication exposing them to the components of the system.[97]

Halliday's analogy between the linguistic system and climate helpfully illustrates the necessity of using a corpus for studying instantiation.[98] He compares the linguistic system to climate and individual texts to weather.[99] Just as one could not justify an opinion about the world's climate solely on the basis of the temperature at one place on

---

[92] Halliday, "Theory to Work," 130; Halliday, "Towards Probabilistic" 44.
[93] Halliday, "Towards Probabilistic," 45.
[94] Halliday, "Spoken," 21; cf. Hasan, "Language and Society," 27.
[95] Halliday, "System and Instance," 82.
[96] Halliday, "Spoken," 21.
[97] Cf. Halliday, "Driving Force," 181; Halliday, "Towards Probabilistic," 44–45.
[98] Hasan, "Language and Society," 27.
[99] Cf. Halliday, "Towards Probabilistic," 45.

particular day, only a large number of instances will reveal patterns indicating the nature of the underlying system.[100]

From this reasoning process SFL develops a graded scale from what a language user means on a particular occasion (a text) to what meanings the language makes available (the system).[101] In between these extremes lie the intermediate categories of text type and register; these are not separate phenomena but instead two complementary ways of investigating the same phenomenon.[102]

The extreme case of modeling the entirety of a language is not the only case in which the cline of instantiation applies. Intermediate between the instance pole (text) and the potential pole (language system) fall the categories of instance type and subpotential (or text type and register, respectively). By definition, texts belonging to the same type exhibit less variation among themselves; less variation means a researcher needs fewer instances to account for that variation. I will develop this idea further in the section on probabilistic grammar below.

Stratum

Turning to the entailments of SFL as a functional model, its concern for the relationship between language and social structure means that it must account for how extralinguistic entities affect language use. The fact that language is functionally related to the situation makes it easier to understand.[103] SFL models this relationship via the theoretical construct of stratum.

[100] Halliday, "System and Instance," 77.
[101] Cf. Matthiessen et al., *Key Terms*, 121–25.
[102] Cf. Halliday, "Towards Probabilistic," 43–44.
[103] Cf. Halliday, *Social Semiotic*, 18.

With regard to the stratal dimension, most SFL work postulates three linguistic strata (phonology/graphology, lexicogrammar, and semantics) and one extralinguistic one (context), for a total of four; Martin and those who follow his genre model subdivide context into multiple strata.[104] The three linguistic strata further subdivide into two planes: semantics and lexicogrammar belong to the content plane, and phonology (for spoken language) or graphology (for written language) belongs to the expression plane.[105] Semantics is the level of meaning, lexicogrammar is the level of wording, and the expression plane is the level of sounding or spelling.[106] Variation in the expression plane has to do with particular language users' eccentricities, rather than any factor that might pertain to genre, so my analysis focuses on context and the content plane.[107]

Halliday's notion of context may be confusing to readers who are used to the meaning normally attached to that term within biblical studies. Rather than using the term context to refer to other parts of the text under consideration, Halliday follows his teacher J. R. Firth in adopting Malinowski's use of context to refer to two ways in which a text relates to the extralinguistic world.[108] The more general of these, the context of culture, accounts for aspects of the language users' cultural background that affect the exchange of meanings during communication.[109] The other sort of context, the context of situation, refers to aspects of the environment "from which the things which are said derive their

---

[104] Hasan, "Language and Society," 18; cf. Matthiessen et al., *Key Terms*, 205.

[105] Halliday, "Gloosy Ganoderm," 107; cf. Matthiessen et al., *Key Terms*, 38.

[106] Halliday, *Social Semiotic*, 20.

[107] Cf. Halliday and Hasan, *Language*, 41; Hasan, "Language and Society," 8–9; Matthiessen et al., *Key Terms*, 94.

[108] Halliday and Hasan, *Language*, 5–8; cf. Halliday, "Machine Translation," 24; Hasan, "Meaning," 55–56.

[109] Halliday and Hasan, *Language*, 6–7.

meaning."[110] I now turn to discussing both of these before moving to the next stratum below.

Whereas the context of situation is "the immediate environment," the context of culture is "a broader background against which the text has to be interpreted."[111] As mentioned above, culture itself is a semiotic construct; it can be further described as "an integrated body of the total set of meanings available to a community" or as the community's "total semiotic potential."[112] This semiotic potential involves not only language, i.e. "ways of saying," but also activities and statuses, i.e. "ways of doing" and "ways of being."[113] The signs of which culture is constituted remain for the most part the same across cultures, but the meaning attached to those signs differ.[114] Hasan gives the example of a term in her native language that refers to the third day after the death of someone: people everywhere die and, consequently, there is a third day after that occurrence, but this day only takes on cultural significance within one particular culture.[115] The semiotic potential of a culture determines the aspects of a material situation on which members of that culture will place stress.[116] The representation of these significant aspects within language constitutes the "semantic potential," which SFL models in terms of the set of possible values for field, tenor, and mode.[117] One particular subset of the semantic potential, what Hasan characterizes as the "genre-specific semantic potential," exists within language as one particular set of values for field, tenor, and

[110] Halliday, *Social Semiotic*, 28.
[111] Halliday and Hasan, *Language*, 46.
[112] Halliday and Hasan, *Language*, 99.
[113] Halliday and Hasan, *Language*, 99.
[114] Halliday and Hasan, *Language*, 100.
[115] Halliday and Hasan, *Language*, 100.
[116] Halliday and Hasan, *Language*, 101.
[117] Halliday and Hasan, *Language*, 101–2.

mode.[118] The genre-specific semantic potential is one particular contextual configuration, which is a type of situation that serves to account for many actual situations.[119]

Malinowski developed his idea of context as a result of his work as an anthropologist among the peoples of the Tobriand Islands in the South Pacific.[120] In the course of his fieldwork he transcribed a number of texts in the islanders' language that he wanted to use as evidence for his ideas about their culture, but he had difficulty making the texts understandable for English-speaking readers as a result of the huge gap between the cultures involved. He found that he had to describe the extralinguistic environment surrounding the transcribed text in order for the text to be understood, and he termed this the context of situation. Even texts that seem separate from their extralinguistic environments, such as narratives of other times and places, often serve a purpose in their situation, like maintaining the groups for whom the narrative is recounted.

"Situation" is far too broad of a term to be useful as a theoretical term without further specification.[121] Specifically, the context of situation is those aspects of the environment that significantly affect the language event.[122] The context of situation "refers to that part of reality which is filtered through the interactants' focus upon some aspect of their environment in performing some social activity."[123] With regard to my project in particular, one of the features common to all the candidate analogies for Mark is that they are "monologic" texts with no turn-taking; this facilitates studying the

---

[118] Halliday and Hasan, *Language*, 102–4.
[119] Halliday and Hasan, *Language*, 102.
[120] Halliday and Hasan, *Language*, 6–7.
[121] Hasan, "Language and Society," 7.
[122] Halliday, *Social Semiotic*, 29; cf. Hasan, "Language and Society," 7.
[123] Hasan, "What's Going On," 79.

situational context because such a text "is so constructed here as to verbally encapsulate its own context maximally."[124]

Exactly what sort of aspects may be relevant depends on the nature of the language event in question: the context of situation for a child using language is often very "concrete and immediate" because they are trying to accomplish something that directly relates to the material environment around them, while that of two experts involved in shop talk "may be quite abstract and remote" because their discussion assumes quite a bit of shared knowledge but might not involve the objects around them at the time of speaking at all.[125] Hasan notes that the overlap between material setting and context of situation often depends on the "role that language plays in the social process": contexts in which language accomplishes the social activity (e.g. a seminar) tend to not have their contexts of situation depending on the material setting, whereas elements of the material setting are often included in the context of situation in contexts where "the role of language is subsidiary."[126]

This concept of situational context has evolved greatly from Malinowski's original formulation of it.[127] Malinowski's original formulation was related to the specific texts on which his reconstruction of the islanders' culture depended, and—as a result—he simply expounded the features that were relevant to those specific instances, rather than generalizing what sorts of features would normally be relevant.[128] Systematizing these features to make the idea of the context of situation useful for linguistic theory fell to one

---

[124] Hasan, "What's Going On," 90.
[125] Halliday, *Social Semiotic*, 29; cf. Halliday and Hasan, *Language*, 99–100.
[126] Hasan, "What's Going On," 81–83.
[127] Hasan, "What's Going On," 78.
[128] Halliday and Hasan, *Language*, 8.

of his younger colleagues at the University of London, the aforementioned J. R. Firth.[129] Firth categorized relevant aspects of the environment under four heads: (1) the participants, (2) the action (whether verbal or non-verbal), (3) the effects, and (4) other.[130] Other scholars have put forward a number of other such categorizations.[131]

Halliday builds his own proposal on the fact that communication often occurs successfully: he suggests that information about the environment surrounding the use of language (i.e. the context of situation) improves the inferences a recipient makes about a speaker or writer's intended meaning.[132] This being the case, a useful description of the context of situation "links it up with the expectation of what others are likely to say."[133] Noting the difficulties with tying previous theories of situational context to texts, Halliday emphasizes the advantages of categorizing elements of a situation that "determine" a text under three headings (namely field, tenor, and mode) that relate directly to the nature of the linguistic system.[134] Together the field, tenor, and mode constitute a text's register

The meanings associated with each type of situational context—a "register" to use SFL parlance—form "a package, so to speak, of things that typically go together in the culture" rather than being "a random jumble of features."[135] The ordered relationship between a text, its situation, and the culture from which it arises is part of what allows language to function as a medium for communication: listeners or readers are aware of the cultural conventions governing their interaction and the relevant aspects of the

---

[129] Halliday and Hasan, *Language*, 8; cf. Hasan, "Language and Society," 7–8.
[130] Halliday and Hasan, *Language*, 8.
[131] Halliday and Hasan, *Language*, 9.
[132] E.g. Halliday and Hasan, *Language*, 45; cf. Halliday, *Social Semiotic*, 62.
[133] Halliday and Hasan, *Language*, 10.
[134] Halliday, *Social Semiotic*, 61–63.
[135] Halliday and Hasan, *Language*, 46.

situation around them, and they exploit this knowledge to correctly interpret the texts they receive.[136] People gain this ability as a result of previous exposure to texts.[137]

Between context and the linguistic system proper lies the semantic stratum, the level of meaning, which consists of “the set of strategies for construing, enacting and presenting non-language as language.”[138] In other words, the semantic stratum serves as the middleman between language “and some higher-order symbolic system,” namely the social system.[139] Recognizing that the semiotic nature of language implied that all strata of language contributed to the creation of meaning, Halliday’s mentor J. R. Firth did not include a stratum specifically devoted to semantics.[140] Halliday’s initial forays into systemic theory followed Firth on this point, but subsequent research showed the need for specifically treating resources for meaning potential within a language separately from the formal potential of language.[141] Specifically, without a specific semantic stratum “there were no terms for talking about meaning” systemically (i.e. paradigmatically) in the same way that one could develop system networks for a language’s formal potential.[142]

The internal makeup of the semantic stratum is still a matter of some debate. One potential perspective is considering the semantic stratum in terms of its relation to the contextual stratum—i.e. viewing it “from above.” The other prospective understanding of

---

[136] Halliday and Hasan, *Language*, 47.
[137] Halliday and Hasan, *Language*, 47.
[138] Matthiessen et al., *Key Terms*, 189.
[139] Halliday, *Social Semiotic*, 79.
[140] Halliday, “Pinpointing,” 143; cf. Hasan, “Language and Society,” 14–15.
[141] Hasan, “Language and Society,” 15–16.
[142] Hasan, “Language and Society,” 16.

semantics is to describe it in terms of its relationship to the stratum of lexicogrammar—i.e. the view "from below."[143]

Viewing semantics from above leverages the metafunctional nature of language: three metafunctions means that context relates to meaning in three ways. Ideational systems relate to field variables, interpersonal systems relate to tenor variables, and textual systems relate to mode variables.[144] Theoretical discussion within SFL offers at least two means for describing the semantic stratum from above—Hasan's notion of a "contextual configuration" and Halliday's concept of "register." These concepts are clearly related because both Halliday and Hasan describe their respective theoretical constructs in terms of the contextual variables field, tenor, and mode.[145]

Hasan uses the notion of contextual configurations for predicting the identity and order of semantic units within a structure of a text.[146] After discussing a range of instances of a particular contextual configuration a general structure that can account for all the aspects of the texts emerges and Hasan calls this—at least in her earlier work—a "generic structure potential."[147] The specification of specific elements of a contextual configuration continues only to the degree necessary to motivate the GSP.[148] Hasan refers to genre as the "verbal expression" of a contextual configuration so that the structure of a text belonging to one genre is the realization of a particular GSP.[149]

---

[143] The unusual term "lexicogrammar" stems from Halliday's conviction (e.g. "Theory at Work," 134) that grammar and lexis are not two different things.

[144] Matthiessen et al., *Key Terms*, 189.

[145] Halliday and Hasan, *Language*, 55.

[146] Halliday and Hasan, *Language*, 56–63.

[147] Halliday and Hasan, *Language*, 63–66. In some of her more recent works, Hasan (e.g. "Language and Society, 20; Hasan, "What's Going On," 94n4) explains GSP as "generalized structure potential."

[148] Halliday and Hasan, *Language*, 105.

[149] Halliday and Hasan, *Language*, 108.

Hasan's perspective would seem to entail that determining a work's genre rests entirely on tracing the progress of the work's semantic structure: the only way in which two works cannot vary and still belong to the same genre is the presence of the obligatory elements of the genre's GSP.[150] In terms of relating a GSP to then notion of register, a GSP describes a contextual configuration, which is defined in terms of field, tenor, and mode like a register, but it lacks the concrete realizations in text that registers have.[151] Given these two alternative descriptions of the semantic stratum, I will stick with register, since its concrete realizational patterns allow for more verifiable analysis.

Viewing semantics from below leverages the idea of rank: a hierarchical scale seems to characterize the semantic stratum just like lexicogrammar combines words into groups and groups into clauses.[152] The largest semantic unit is the text, which in this sense refers to "language operating in context."[153] Typically, texts consist of parasemes—i.e. "rhetorical paragraphs"—which in turn consist of sequences.[154] Sequences consist of three "semantic units" corresponding to the three metafunctions: move (interpersonal), figure (ideational), and message (textual).[155] Using lexicogrammar as a point of entry into semantics entails correlating the two rank scales. Sequences relate to clause complexes, and the semantic units typically relate to the clause.[156]

---

[150] Halliday and Hasan, *Language*, 108.

[151] Cf. Halliday and Hasan, *Language*, 68.

[152] The rank scale of the semantic stratum differs from the rank scale of the lexicogrammatical stratum discussed below in that it lacks formal criteria to mark off units (cf. Halliday and Hasan, *Language*, 10).

[153] Matthiessen et al., *Key Terms*, 218; cf. Halliday and Hasan, *Language*, 10.

[154] Matthiessen et al., *Key Terms*, 190.

[155] Matthiessen et al., *Key Terms*, 190.

[156] Matthiessen et al., *Key Terms*, 190.

Lexicogrammar, the next stratum below semantics, is where the abstract meanings of a text take on flesh in the structures of a language.[157] Halliday refers to this stratum as "the purely internal level of organization, the core of the linguistic system."[158] The token of the meaning (i.e. the semantics) of a text is the wording, i.e. the set of words that instantiates the text. In terms of Halliday's description of text to which I referred earlier, a given wording enshrines a particular "current of selection" whose meaning depends on "its paradigmatic environment."[159]

Metafunction

Adopting a functional view of language means focusing on "what the speaker, child or adult, can do with" language and describing "its internal organization and patterning, in terms of the functions that it has evolved to serve."[160] This perspective necessitates determining what these functions are.[161] SFL conceives of two major purposes that languages serve: (1) representing entities and occurrences in the world around the language user and (2) enacting and maintaining social relationships. In addition, if communication is to occur, language must accomplish these goals in such a way that the recipient can retrieve the information.[162] SFL categorizes the meaning-making resources a language uses to accomplish these three goals into three groups, called "metafunctions."[163] Representational resources fall under the ideational metafunction,

---

[157] Cf. Halliday, "Gloosy Ganoderm," 108.
[158] Halliday, *Social Semiotic*, 43.
[159] Halliday, "Text,"48.
[160] Halliday, *Social Semiotic*, 17; cf. Halliday and Hasan, *Language*, 44.
[161] Halliday, *Grammar*, xv; Halliday, "Theory to Work," 135; cf. Matthiessen et al., *Key Terms*, 138.
[162] Halliday, "Theory to Work," 135–36; cf. Thompson, *Introducing*, 30.
[163] Halliday, "Gloosy Ganoderm," 107–8; cf. Matthiessen et al., *Key Terms*, 38.

and resources for social interaction fall under the interpersonal metafunction.[164] The textual metafunction covers resources for accomplishing the third goal by making language "relevant to its environment, as distinct from decontextualized language like words listed in a dictionary or sentences in a grammar book."[165] Every complete text means in these three ways.[166]

Both systemic and structural evidence support dividing semiotic resources into these three categories. In terms of system, description of English's lexicogrammar reveals three "clusters of systems with strong interconnections within each cluster, but weak associations outside the cluster," indicating three "discrete" categories within English lexicogrammar.[167] Additionally, the contribution each of these metafunctions makes to the overall structure of a clause brackets clausal constituents differently, whereas one would expect overlap if any of the metafunctions were redundant.[168]

The metafunctions operate at both the semantic and lexicogrammatical strata, allowing them to organize connections observed between these strata; they "explain the internal nature of language in such a way as to relate it to its external environment."[169] If one were to extend the diagram Matthiessen and his coauthors use to describe the relationship between strata to account for the metafunctional dimension, then three wedge-shaped sections would cut across the concentric circles representing strata; each wedge would correspond to a metafunction.[170] A "function-stratification matrix" displays

---

[164] Cf. Halliday, "Language in Relation to Fuzzy Logic," 200–01.
[165] Halliday, "Text," 29; cf. Halliday and Hasan, *Language*, 45.
[166] Halliday, "Gloosy Ganoderm," 108; Halliday, *Social Semiotic*, 56; Halliday, "Theory to Work," 135.
[167] Webster, "Introduction," 5–6; cf. Halliday, *Social Semiotic*, 46.
[168] Halliday, "Theory to Work," 135; cf. Webster, "Introduction," 6; Thompson, *Introducing*, 34.
[169] Halliday, *Social Semiotic*, 48; Hasan, "Language and Society," 22; cf. Halliday, "Towards Probabilistic," 44.
[170] Matthiessen, *Key Terms*, 207; cf. Halliday, *Social Semiotic*, 187–89.

what systems operate in each of the metafunctions at these strata.[171] Although the concept of such a matrix is theoretical—and, as such, language-independent—the systems appearing in the cells are language-dependent, so the matrix for Hellenistic Greek may not be identical to the English one.[172] Nonetheless, the English function-stratification matrix provides the starting point for generating one for Hellenistic Greek, so I will return to it in due course.

Summary

To summarize the import of this whole discussion of SFL theory for my particular project, my task is two-fold: I must describe the total potential at each level, and then I must relate the levels to each other.[173] A semiotic dimension undergirds both stages. Describing the total potential of a level involves the instantiation; relating the levels to each other involves the stratal dimension.

I should perhaps say something here about the fact that SFL has predominately been used to study English.[174] This limitation is not nearly as serious as it might seem to be at first glance because SFL demarcates clearly between parts of the theory that apply to a language in particular and which parts apply to language in general. SFL's core concepts, or theory, are posited as language universals.[175] Descriptions, however, differ because for a paradigmatically-oriented model "describing something *consists in* relating it to everything else," and different languages will have different options and, therefore,

[171] Matthiessen, *Key Terms*, 104.
[172] Matthiessen, *Key Terms*, 104.
[173] Halliday, *Social Semiotic*, 42–43.
[174] Cf. Porter, "Systemic Functional Linguistics," 17–20.
[175] Halliday, *Social Semiotic*, 21.

different relationships between those options.[176] Similarly, analyses are tied to descriptions, so they are language-dependent.[177]

Halliday illustrates the distinction between these elements by referring to the metafunctions: the organization of semantics and lexicogrammar into ideational, interpersonal, and textual components is a theoretical concept whose validity is posited for all languages, but Halliday makes no guarantee that any given language will have a thematic system or (if there is one) that the options within it will be the same as those of the English thematic system.[178] The Hellenistic Greek system of CAUSALITY exemplifies this in that English has a similar system, but the Greek one has an extra term, namely the middle voice represented in Porter and O'Donnell's system network as +ergative.[179]

Halliday mentions that his picture of English has informed how other researchers have described other languages by a two-fold process: they use the categories introduced by the English description and attempt to find them in the language, and then they consider whether or not they would have come to the same conclusion without having used English as a heuristic.[180]

A bonus consideration in favor of choosing SFL as the linguistic model for this project is that a well-developed body of work shows it to be productive for analyzing Hellenistic Greek. This body of work concerns a wide variety of issues including the discourse function of connectives in Matthew, the way Luke signals the importance of characters in Acts, and verbal semantics.[181]

---

[176] Halliday, *Grammar*, xxvii.
[177] Halliday, *Grammar*, xxii.
[178] Halliday, *Grammar*, xxxiv.
[179] Porter and O'Donnell, "Greek Verbal Network," 40.
[180] Halliday, *Grammar*, xxxiv.
[181] E.g. Black, *Sentence Conjunctions*, passim; Martín-Ascensío, *Transitivity*, passim; Porter and O'Donnell, "Greek Verbal Network," 3–41.

Instances of language use vary from each other for a variety of reasons, many of which are arbitrary and, therefore, not subject to scientific scrutiny. Two specific sorts of variation, however, are systematic enough to be studied, namely dialect and register. Both of these "*actively symbolize*" the variety that characterizes social structure, but they do so in different ways because they correlate with different aspects of extralinguistic context.[182] Dialect variation relates to diversity of social structure, and register variation relates to diversity of social activities.[183] Another way of characterizing the difference is that "dialects are saying the same thing in different ways, whereas registers are saying different things."[184]

I focus here on register variation because it better correlates with the task of characterizing Mark's genre: choosing a different genre is saying a different thing, not expressing the same thing differently.[185] The "very simple and very powerful" concept of register accounts for varieties of language that correspond to varied contexts.[186] This term refers to collections of instances that exhibit similarities because they arose from similar extralinguistic contexts that consequently exerted similar pressures on semantic and lexicogrammatical choices.[187]

Halliday calls attention to the work of Bernstein and Labov as paving the way towards revealing the "elusive relation between language and social structure."[188] Labov

---

[182] Halliday, *Social Semiotic*, 3.

[183] Halliday, *Social Semiotic*, 2.

[184] Halliday and Hasan, *Language*, 41.

[185] Admittedly, this formulation does somewhat oversimplify the case since sometimes a particular register necessitates a particular dialect (Halliday, *Social Semiotic*, 2–3; cf. Halliday and Hasan, *Language*, 42).

[186] Halliday, *Social Semiotic*, 31; cf. Halliday and Hasan, *Language*, 38.

[187] Halliday, Gloosy Ganoderm," 109; Land, "Varieties," 243–60; Porter, "Dialect and Register," 197–207; Porter, "Sociolinguistics," 124–27.

[188] Halliday, *Social Semiotic*, 2.

showed that it is “normal for variations to occur systematically.”[189] In contrast to ethnomethodological approaches to language study, register study within a SFL framework does not focus on ad hoc analysis of individual instances; rather, it looks for the regularities that characterize language use in particular sorts of contexts and ascertains the principles that create these regularities.[190] Remembering that these regularities are precisely that (and not rules) is important: “typical behaviour is not invariant behaviour even within the context of the same culture.”[191] Hasan, drawing on some early work by Halliday and collaborators, points out two ways to find these regularities and principles: (1) one may look at the “distinctive character of the social situation” characteristic of the register or (2) one may look for the “demands a category of register . . . would make on the system of language.”[192] These perspectives are another way of viewing the approaches to semantics, i.e. “from above” and “from below,” respectively, which makes sense because registers are clusters of meaning, specifically the “configurations of meanings that are typically exchanged—that are ‘at risk’, so to speak—under given conditions of use.”[193]

Approaching register from above involves relating the extralinguistic context to the meanings associated with it.[194] Since culture is itself a system of meanings, this essentially means connecting the meanings of the social system to those of the linguistic system.[195] According to Halliday, language encodes three sorts of features that describe

[189] Hasan, “What’s Going On,” 86.
[190] Hasan, “Language and Society,” 7; cf. Hasan, “What’s Going On,” 87–88.
[191] Hasan, “What’s Going On,” 87.
[192] Hasan, “Language and Society,” 7.
[193] Halliday, *Social Semiotic*, 185.
[194] Cf. Porter, “Systemic Functional Linguistics,” 20.
[195] Halliday, *Social Semiotic*, 189.

the situation in which the language use occurred; he calls these field, tenor, and mode.[196] These components constitute “certain systematic norms governing the particulars of the text.”[197] Field refers to what is going on in a particular situation; it relates to the idea of topic.[198] Tenor describes the “cluster of socially significant relationships” between participants in the discourse.[199] Mode differs from the other two in that it is primarily concerned with the means by which the other two components are expressed.[200]

As mentioned above, the linguistic system also contains three independent components (the metafunctions). The existence of three situational components and three metafunctions suggests that there should be a metafunction within language corresponding to each component of the extralinguistic situation, and, indeed, this is the case. Field corresponds to ideational choices, tenor corresponds to interpersonal choices, and mode corresponds to textual choices.[201]

This is something most people can do intuitively during the course of their social interactions: they can, for example, show up to a gathering and “size up” the social activity that is taking place, the role relationships among the participants, and the means by which communication is occurring. As a result, they can understand how to integrate themselves into the exchange of meaning that is occurring.[202] In point of fact, communication only successfully occurs in many contexts because of this process; often, only a small part of the meanings one needs to pick up are explicitly realized in wording:

---

[196] Halliday and Hasan, *Language*, 12; Halliday, *Social Semiotic*, 32; Land, “Varieties,” 251–52.
[197] Halliday, *Social Semiotic*, 62.
[198] Halliday, “Text,” 55–56; cf. Porter, “Ideational,” 147–153.
[199] Halliday and Hasan, *Language*, 12; Halliday, “Text,” 56.
[200] Halliday, “Text,” 57.
[201] Halliday, *Social Semiotic*, 63; Halliday and Hasan, *Language*, 25.
[202] Halliday, *Social Semiotic*, 189.

"we succeed in the exchange of meanings because we have access to the semiotic structure of the situation from other sources."[203]

The approach from below, via lexicogrammar, has dominated SFL register study.[204] This involves relating the content plane of language (semantics/grammar) to the context of situation.[205] Since SFL-based descriptions already systematically relate observable features of texts (i.e. the expression plane) to the content plane, relating the content plane to the context of situation provides a justifiable bridge from a text to its context of situation.[206] SFL presupposes that language is a system and that all language exists within—and relates to—a context of situation, so the question in which such study is interested is what factors of a situation correlate with what selections from the linguistic system.[207]

Recipients of communication improve their interpretations of the communication by using their knowledge of the situation to fill in blanks; register study is a promising avenue for reversing this process by reconstructing the situation from which a text arose on the basis of the text itself.[208] One should note, however, that investigation of a text will not allow the reconstruction of all the details of the extralinguistic context from which the text sprang; the text will only reveal the aspects of the situation that have affected the text's wording.[209] Thus, a situation type or register reconstructs the relevant context, i.e. that part of the material setting that is linguistically represented.[210] Thus,

---

[203] Halliday, *Social Semiotic*, 189.
[204] Hasan, "Language and Society," 9.
[205] Halliday and Hasan, *Language*, 34–35.
[206] Halliday and Hasan, *Language*, 34–36.
[207] Halliday, *Social Semiotic*, 32.
[208] Halliday and Hasan, *Language*, 36–38.
[209] Hasan, "Language and Society," 11.
[210] Hasan, "Language and Society," 12.

Hasan argues, accurately reconstructing the context of situation would allow an investigator "to explain why certain things have been said or written on this particular occasion, and what else might have been said or written that was not."[211]

Linking this last quotation from Hasan with Halliday's definition that "text is meaning and meaning is choice…in its paradigmatic environment of what **might have been** meant (but was not" shows where register falls in terms of stratification and, thus, how it relates to observable features of a text. Hasan describes reconstructing a context of situation in terms similar to Halliday's definition of the meaning of a text; this coheres with the definition of register found in one of their joint publications: register "can be defined as a configuration of meanings that are typically associated with a particular situational configuration."[212] In other words, registers are patterns of meaning that together signify a particular situation. As discussed in the introduction to SFL above, strata relate to each other as "realizations," or significations of higher strata. The stratum below context is the semantic stratum, i.e. the stratum of meaning, so seeing a pattern of meanings as signifying the context makes sense within the context of the theory. Pursuing this farther, SFL theory specifies that lexicogrammar realizes the semantic stratum, so a register should have patterns of lexicogrammar that characterize it, and this does in fact seem to be the case.[213] These realizing patterns take many forms, but the commonality is that they involve "formal linguistic features" that one can observe in text.[214] Thus, register theory provides a concrete bridge from observable features of a text to aspects of the context from which it arose.

---

[211] Halliday and Hasan, *Language*, 46.
[212] Halliday and Hasan, *Language*, 38–39.
[213] Halliday and Hasan, *Language*, 39; cf. Halliday, "Text," 57–58.
[214] Porter, "Ideational," 148; cf. Halliday and Hasan, *Language*, 39.

Having described the warrant for linking observable features of a text to the situation that generated it, the next task is to determine which features of a text relate to which features of a situation. The values of these three variables are what differentiate registers from one another: if the field, tenor, or mode differs between two texts, then those texts belong to separate registers, even if the situations behind the texts would seem to be the same on extralinguistic grounds.[215]

This correspondence allows one to move up the hierarchy of realization mentioned earlier: given a description that specifies the ideational, interpersonal, and textual components of the language in which a text is written, one can catalogue the ideational, interpersonal, and textual choices an author made in constructing the text and thus describe the register of the text, which will in turn bear witness to the text's context of situation.[216] In short, this will be the avenue I take for my project: I will catalogue the choices in three systems (one from each metafunction and, therefore, representing all three components of a situation); I will return to this idea further as I lay out my procedure.

The relationship between genre and register is somewhat unclear within SFL theoretical discussion. The evolution of Hasan's notion of GSP emblemizes this. In some of her work, particularly early on, she glosses this acronym as "generic structure potential." Later on, however, she notes that she would prefer to gloss it as "generalized structure potential," removing the reference to genre and linking it with register: it "is designed to represent the possible range of the structure available within a given

---

[215] Halliday, *Social Semiotic*, 31.
[216] Halliday, *Social Semiotic*, 63; cf. Hasan, "What's Going On," 81.

register."[217] Thompson suggests that genre is "register plus purpose": genres appropriate "the resources of a register (or more than one register) in particular patterns to achieve certain communicative goals."[218]

For Halliday, explanations for the similarities exhibited by the constituent instances differentiate a text-type from a register; the register is abstracted further towards the system pole of the cline of instantiation, operating as a sub-system.[219]

This notion has proven to be "one of the most productive" contributions of SFL to the study of the New Testament.[220] Advantages of SFL-based register study for studying ancient texts include: (1) it "bridges the gap" between investigating people as language users and investigating texts as the product of such use, (2) it bridges the gap between form and meaning, which is strongly to be desired with regard to a language where we only have access to written remains, and (3) it forces an interpreter to come to terms with all the various facets of language, rather than simply cherry-picking features that jump out at him or her.[221]

Linguists working within a SFL framework have produced a number of register analyses of texts—or groups of texts—drawn from a variety of genres. However, such work has not represented all genres equally: some genres (like science) have received disproportionate attention, while other genres (such as history) have received very little.[222] Contrariwise, discourse analysis of historical texts has often used a top-down

---

[217] Hasan, "What's Going On," 94n4. This note, which Hasan added in the process of preparing this article for inclusion in her collected works, clearly demonstrates the development: the sentence in the article to which the note is attached connects the GSP with genre (Hasan, "What's Going On," 84).

[218] Thompson, *Introducing*, 43.

[219] Halliday, "Theory to Work," 131.

[220] Porter, "Mark 13," 219.

[221] Porter, "Sociolinguistics," 126.

[222] Martin and Wodak, "Introduction," 1.

approach, rather than attending to in-depth study of how a historian deploys the lexicogrammatical resources of a particular language to construct his or her picture of the past.[223]

The Surveying: Grammatical Probability

Halliday's work on grammar as a probabilistic system stands as one of his most fundamental contributions to the study of language.[224] A conviction that language is inherently probabilistic has characterized Halliday's perspective throughout his career, ever since his early work on Chinese grammar.[225] Of particular relevance to this study is the fact that it can put grammatical data into a quantitative form that allows for statistical comparison. This section describes the mathematical justification for grammatical probability and the systems whose probabilities I intend to discover.

Justification for Probabilistic Grammatics[226]

The notion of probability within linguistic description lies "between the certainty of system and the uncertainty of instance."[227] As described above, SFL recognizes a graded scale from what a language user means on a particular occasion (an instance) to what he or she could mean given the resources of the language in which he or she is communicating (the system).

---

[223] Martin and Wodak, "Introduction," 1–2.

[224] E.g. Halliday, "Towards Probabilistic,"42–62; Halliday, "Corpus Studies," 63–75; Halliday, "System and Instance, 76–92; Halliday, "Quantitative Studies," 130–56. Applications of the model include a study of polarity and tense in English (Halliday and James, "Polarity and Primary Tense," 93–129) and a study of the verbal system of Hellenistic Greek (Porter and O'Donnell, "Greek Verbal Network," 3–41).

[225] Halliday, "Corpus Studies," 64.

[226] Halliday (e.g. "System and Instance," 76) notes the confusion inherent in using the word grammar to refer to both what is being studied and the discipline devoted to such study. He introduced the term grammatics to refer to the science of studying grammar.

[227] Hasan, "Language and Society," 27.

Halliday's notion of grammatical probability rests on the interaction between instantiation and stratification.[228] Since text instantiates the system, texts bear witness to all the system's features, including the relative probabilities of terms within the system.[229] Frequency of forms known to be realizations of particular terms in a system instantiate the probabilities attached to the associated terms.[230] In other words, the grammatical probability of a systemic feature is the number of times it was chosen (i.e. its frequency) divided by the total selections in that system. Thus, grammatical probability provides a clear methodology for proceeding from the concrete forms in the text to less material aspects of the text.

The text's genre is one such immaterial aspect. In at least some of his work Halliday characterizes genre as somewhat equivalent to situation type.[231] In terms of the intersection between instantiation and stratification situation types exist as subpotentials in the contextual stratum as an intermediate point between the situation of an individual text and culture as the set of all possible situations for texts.[232] Registers and the associated concept of text types (discussed above) serve as the subpotential of both the semantic and lexicogrammatical strata.[233] Given the realizational relationship between strata, this means that registers are a symptom of genre. My task in ascertaining the best analogical genre for the Gospel of Mark, then, is to describe the registers of candidate analogies.

---

228 Halliday, "Towards Probabilistic," 61.
229 Halliday, "Corpus Studies," 64.
230 Halliday, "Towards Probabilistic," 45.
231 Matthiessen et al., *Key Terms*, 107.
232 Matthiessen et al., *Key Terms*, 125.
233 Matthiessen et al., *Key Terms*, 125.

Grammatical probability is supremely suited to describing register variation, for such variation "can be defined as the skewing of (some of) these probabilities, in the environment of some specific configuration of field, tenor, and mode."[234] Another way of phrasing this is to say "genre may be nothing more than a register's specific configuration of metafunctions."[235] Furthermore, Halliday characterizes register elsewhere as "a form of prediction," and probabilities are simply quantified predictions.[236]

Elements of the context (i.e. field tenor, and mode) affect which meanings—and, consequently, which forms expressing those meanings—are appropriate in that context.[237] Sometimes the pressure contextual concerns exert on typical systemic probabilities is quite obvious because the register's context habitually calls for choices that are normally infrequent (e.g. the future in weather forecasting or the imperative mood in an instruction manual).[238] However, the "resetting of probabilities" that serves as an indication of register variation need not be as drastic as these examples in order to be significant.[239] In fact, most register variation "is more subtle and complex."[240]

The notion of register variation as systematically skewed probabilities implies that a norm characteristic of the language as a whole exists; otherwise there would be nothing to skew. Overall, the grammatical probabilities within a system remain stable. Each

---

[234] Halliday, "Towards Probabilistic," 60; cf. Halliday, "System and Instance," 84–85.
[235] Pitts and Tyra, "Exploring Linguistic Variation," 262.
[236] Halliday, *Social Semiotic*, 32.
[237] Halliday, "Towards Probabilistic," 52. Another factor that affects probabilities is code variation, but this sort of variation lies outside my study. I mention it here in passing because recognizing that code represents a possible extraneous factor in my analysis led me to restrict my corpus to Jewish sources in order to minimize the effect of code.
[238] Halliday, "Text," 55; Land, "Varieties," 251–52; Porter, "Dialect and Register," 200.
[239] Halliday, "Information and Meaning," 57.
[240] Halliday, "Information and Meaning," 58. Elsewhere (e.g. Halliday and Hasan, *Language*, 39–40) he refers to registers with obvious realizational patterns as "closed" and ones with subtle realization as "open."

instance affects them, but such disturbances "are too small to be taken account of, and mostly cancel each other out."[241] In fact, the probabilities typically attached to various terms in a system are part of the systemic knowledge that allows a language to function as a system of communication.[242]

Halliday notes that some scholars have objected to his notion of grammatical probability because every text is in a particular register; Halliday admits that this is the case but denies that it constitutes a valid objection to his theory.[243] Rather, he responds, the presence of register variation simply necessitates considering a variety of texts in order to produce a reliable quantitative picture of a language's grammar.[244] Returning to the analogy between instantiation and climate, Halliday notes that a number of factors influence temperature observations, but this does not disallow climatology as a science.[245] Similarly, Halliday expresses consternation with the fact that some of the same people who object to the idea of grammatical probability have no objection to discussion of lexical probabilities.[246]

Instead of counting against the idea of grammatical probability, modelling grammar probabilistically is the very basis upon which one can convincingly explain register variation: such variation "can be defined as the skewing of (some of) these probabilities, in the environment of some specific configuration of field, tenor, and mode."[247] This means that comparing the grammatical probabilities exhibited by a text of uncertain register to those exhibited by a corpus of text representing a known register can

---

[241] Halliday, "Towards Probabilistic," 45; cf. Halliday, "Corpus Studies," 67.
[242] Halliday, *Grammar*, xxii; Halliday, *Social Semiotic*, 61; Halliday, "Towards Probabilistic," 51.
[243] Halliday, "Towards Probabilistic," 59.
[244] Halliday, "Towards Probabilistic," 59.
[245] Halliday, "Towards Probabilistic," 59; Halliday, "Corpus Studies," 70.
[246] Halliday, "Corpus Studies," 64–65; Halliday, "Towards Probabilistic," 59–60.
[247] Halliday, "Towards Probabilistic," 60.

indicate whether or not the work in question matches the known register. In the case of my study, this means comparing Mark to a corpus of texts that represent referential narratives, such as histories and biographies, and also to a corpus of texts representing non-referential narratives, such as novelistic works.

The idea of grammatical probability probably seems strange at first, but it follows directly from the paradigmatic focus of SFL.[248] If a system's entry condition is met, there is a 100 percent chance of the language user selecting one of the system's options, which means the sum of the probabilities attached to the options in that system will be 1.00.[249] Most systems consist of sets of binary oppositions.[250] Knowing that probabilities have to sum to 1.00 and that systems are normally binary allows one to determine the probabilities for options within a system that do not have a direct realization in text: the probability of the term without direct realization will equal 1.00 minus the relative frequency of the form that realizes the realized option.

Beyond organizing the grammatical system paradigmatically so that the rules of probability theory apply, a second key component of the qualitative framework that renders quantitative grammar meaningful results from Halliday's concept of lexicogrammar and the difference between the content and expression planes. People have no trouble admitting that certain words are more common than others (e.g. one is more likely to use the word go than the word perambulate), so Halliday's belief that lexis and grammar are two sides of the same phenomenon would entail accepting the relevance

---

[248] Halliday, "Towards Probabilistic," 45; Halliday, "Corpus Studies," 66; O'Donnell, *Corpus*, 31–32; Porter and O'Donnell, "Greek Verbal Network," 11–13; cf. Libby, "Disentangling," 176–77.

[249] Cf. Halliday, "Corpus Studies," 65.

[250] Halliday, "Machine Translation," 23.

of grammatical probability as well.[251] Admittedly, this argument will only convince those whom Halliday has already convinced of the unity between lexis and grammar. Another, perhaps more convincing, avenue to pursue is to recognize an important difference between the content and expression planes. Specifically, Halliday notes that content strata relate to each other systematically, but content strata relate to expression strata arbitrarily.[252] As Halliday notes, this dispenses with Chomsky's dismissive comment regarding the relevance of textual frequencies: the objection that "'*I live in New York* is more frequent than *I live in Dayton, Ohio*'" misses the point because grammatical patterns like the ratio of various tense forms "could not be reduced to accidental effects like the population of American cities."[253] The salient difference between grammatical patterns and the sort of randomness to which Chomsky referred rests precisely in the strata at which they operate: grammatical probability is only admissible for the content strata.[254]

Systems' probabilities tend to fall into two categories: (1) systems where the probabilities of the terms are more or less equal and (2) systems where the probabilities differ by an order of magnitude (0.9/0.1).[255] The dominant term in the second sort of system is normally the so-called "unmarked" term that the author chooses when there is no reason to pick the less common term.[256] Since the marked term normally appears only when there is a special reason to use it, interpreters are well-advised to take note of their use and search for the motivation that caused the author to choose it.

---

[251] Halliday, "Corpus Studies," 64–65.
[252] Halliday, *Social Semiotic*, 44–45; cf. Hasan, "Meaning," 42.
[253] Halliday, "Corpus Studies," 63.
[254] Cf. Hasan, "Language and Society," 21.
[255] Halliday, "Towards Probabilistic," 48; Halliday, "Corpus Studies," 69–70.
[256] Halliday, "Corpus Studies," 68.

Choice of Systems

How does one determine these probabilities, however? The key lies in integrating the cline of instantiation, the idea of register, and the corpus of texts. In SFL, the common concept of genre is related to the theoretical concept of text type or register.[257] The cline of instantiation indicates that a large enough corpus of instances (i.e. texts) will reveal the characteristics of a text-type. Therefore, analyzing the corpus should indicate whether referential or non-referential narratives better match the Gospel of Mark. Corpus analysis consists of tracing the "ongoing current of selections each in its paradigmatic environment," recording at each point where a system is activated in the unfolding of a text what systemic option the author chose.[258] Ideally, one would perform this process for every system in the language in order to create a complete picture, but practicality raises its ugly head again here: Halliday reports that—as of 1994—computational analysis of English included "about 1000 systems," and Hellenistic Greek seems unlikely to have sufficiently fewer systems to make full enumeration practical.[259] The original goal was to trace at least three systems (one from each metafunction) in order to represent field, tenor, and mode.[260]

Since the systems are many and the ones to be analyzed are few, I can be highly selective in which systems I use.[261] Certain system types make probabilistic analysis difficult, so I have avoided these systems. First of all, two sorts of systems would require far larger samples to investigate reliably: systems with a marked term and highly delicate

[257] Matthiessen et al., *Key Terms*, 106–7.
[258] Halliday, "Text," 48.
[259] Halliday, "Systemic Theory," 434.
[260] Porter, "Register in the Greek," 209.
[261] Cf. Smith and Seoane, "Categorizing," 214–17.

systems. Second, some systems—particularly within the textual component—do not have explicit, structural realizations whose frequency I can count in order to form a picture of the underlying system. Others are realized with phonological structures, which from the perspective of one confined to a written corpus might as well be no structural realization at all.[262] However, Hellenistic Greek still offers more than three systems that meet these requirements, and I now turn to describing which of these I will use.

In terms of a system to represent field, the default choice would be transitivity, since the function-stratification matrix for English indicates that transitivity is the major ideational system in lexicogrammar.[263] Indeed, in some of his early work on register, Porter pursued exactly this course.[264] However, transitivity analysis of Hellenistic Greek is in its infancy at best; the degree to which English process types and associated participant types carry over into Greek is unclear.[265] In view of these difficulties, perhaps the safest course is to find another, more formally-based way to describe the phenomena SFL handles under the rubric of transitivity.[266]

One potential solution is to leave the experiential part of the ideational metafunction alone and consider a system from the logical component.[267] This component involves how clauses relate to each other.[268] Logical systems would also seem to be a good choice from the standpoint that they tend to have direct lexical realization

---

[262] Cf. Schneider, "Historical Variation," 71.

[263] Matthiessen, *Key Terms*, 106; cf. Halliday and Webster, *Text Linguistics*, 24–26.

[264] Porter, "Dialect and Register," 207; Porter, "Register in the Greek," 222–25.

[265] Thompson (*Introducing*, 86–126) provides an accessible introduction to English transitivity analysis. The only major application of transitivity analysis to the New Testament of which I am aware is Martín-Ascensío's study of Acts (*Transitivity*, passim), although recently Chris S. Stevens has published some small-scale studies (Stevens, "Objective Grammatical Means," 327–49; Stevens, "Advantages," 63–86).

[266] Cf. Porter, "Ideational," 148–53; Porter, "Mark 13," 230–35.

[267] Cf. Halliday, *Grammar*, 192–251.

[268] Halliday, *Grammar*, 193.

within text, which would greatly speed up the collection of data. Unfortunately, however, logical systems also tend to be recursive and, as such, they tend to have radically skewed probabilities.[269] As a result, if Halliday's prediction that a skew system's probabilities tend to be in the ballpark of .9/.1 holds true, generating a statistically viable picture of the system would take a much larger corpus.[270]

On balance, fixing the problems observed with transitivity analysis seems the best avenue. At its root, transitivity analysis attempts to describe how participants involved in an action relate to that action. The voice system, which Porter and O'Donnell label CAUSALITY, offers a formally-based means of investigating this: although one loses the interpretive depth of considering other experiential elements, such as circumstances, the upside is that one can jettison debatable theoretical constructs like process types.[271] Furthermore, circumstances and other experiential elements will come into play within the system I have chosen for mode. Unfortunately, problems in defining voice forms sufficiently rigorously for meaningful statistical analysis forced me to give up on studying the CAUSALITY system.[272]

Tenor is the semantic component that is easiest to link to a lexicogrammatical system. A number of different systems fall under tenor, including mood, modality, and person.[273] SFL often uses verbal mood to illustrate the interpersonal metafunction in

---

[269] Halliday, "Towards Probabilistic," 46–47.

[270] Halliday, "Towards Probabilistic," 47.

[271] Porter and O'Donnell, ("Greek Verbal Network," 23n81) note the problem resulting from the morphological ambiguity of middle and passive forms in the present, imperfect, perfect, and pluperfect tense-forms.

[272] Essentially, there were too many ambiguous middle/passive forms. While, as Porter suggests, many of the ambiguous forms can be disambiguated on semantic grounds, I was trying to keep my systems grounded in Greek morphology to facilitate the application of statistics. Meaningful statistical analysis needs clear categories (cf., e.g., Jenset and McGillivray, *Quantitative Historical Linguistics*, 15).

[273] Halliday, *Social Semiotic*, 64.

English text.[274] This trend has continued among those who apply the model to Hellenistic Greek.[275] Porter and O'Donnell have provided a system network for what they term the ATTITUDE system of Hellenistic Greek.[276] Thus, both theoretical and practical considerations argue for choosing this system to represent tenor. Unfortunately, while easy to identify and already networked, there was not sufficient variation in the corpus—particularly in the narrative framework—to allow for meaningful statistical analysis.[277]

Mode is the most difficult semantic component for which to find a representative lexicogrammatical system to model probabilistically.[278] Textual resources within a language fall into two categories, namely structural and cohesive.[279] Since cohesive relations are patterns in the use of other systems, rather than full systems in their own right, one cannot generate the system network for them that would be necessary to make modelling them probabilistically possible.[280] Structural relations consist of information systems and thematic systems.[281] Phonology realizes information systems, at least in English, leaving thematic systems as the most promising option for modern researchers of an ancient language to investigate mode probabilistically.[282]

---

[274] E.g. Halliday and Webster, *Text Linguistics*, 8.

[275] Porter, "Dialect and Register," 205; cf. Porter, "Register in the Greek," 218–22; Porter, "Mark 13," 227–29.

[276] Porter and O'Donnell, "Greek Verbal Network," 40.

[277] Schneider, "Historical Variation," 59–60.

[278] Porter, "Dialect and Register," 200-201; cf. Porter, "Register in the Greek," 218–22 and (most recently) Porter, "Mark 13," 221–27, where he focuses on cohesion and information flow.

[279] Halliday, "Text," 30.

[280] Cf. Halliday, "Text," 30.

[281] Halliday, "Text," 30.

[282] Halliday, "Text," 32–36; cf. Halliday, "Theory at Work," 137. More recently, Halliday and Webster (*Text Linguistics*, 124) suggest that information units for written texts "usually correspond to the clause," but sticking with thematic systems seems wise in view of their qualifier and caveat about punctuation altering the situation.

Thematic systems are ways of signaling what information in the clause is being emphasized.[283] In a number of languages, including English and potentially Hellenistic Greek, the order of elements in the clause realizes thematic status: the Theme is the first element of the clause, and the Rheme is everything else.[284] Halliday recognizes three types of theme, corresponding to the three metafunctions.[285] Interpersonal and textual themes "signal *how*" the clause of which it is a part fits in with the rest of the message, whereas the ideational theme shows "*what* is going to be fitted in."[286] I am concerned here with topical (i.e. ideational) theme defined as the first clausal element with a transitivity function.[287] Defining the object of study in this way allows me to focus on explicit clausal constituents, rather than relying "on the phantom presence of various syntactical phenomena": while one cannot know for certain that any particular clause constituent will appear in a given clause, if at least one such component does not appear, there is no clause.[288]

Clausal elements that function in transitivity include processes (i.e. verbs), participants, and circumstances. Each of these elements helps "explain in the most general way how phenomena of the real world are represented as linguistic structures."[289]

System networks provide the necessary qualitative framework to enable quantitative study of grammar as a probabilistic system. Porter and O'Donnell provided system networks for the ideational and interpersonal systems chosen for this study

[283] Halliday, *Grammar*, 39.
[284] Halliday, *Grammar*, 38; Halliday, "Text," 30–32; Pitts, "World Order," 316.
[285] Thompson, *Introducing*, 158–60.
[286] Thompson, *Introducing*, 159.
[287] Black, *Sentence Conjunctions*, 34; Halliday, *Grammar*, 53.
[288] Porter, "Word Order and Clause Structure," 187.
[289] Halliday, *Grammar*, 102.

(CAUSALITY and ATTITUDE, respectively).[290] However, to my knowledge, nobody has published a system network of the Hellenistic Greek thematic systems, so I am on my own to describe ideational theme choice systemically.

Having discussed the factors that go into generating a system network, what does the network for ideational theme look like? Applying the principle of one-to-one correspondence between forms in realized text and paths through the network to the system of ideational theme, whose most delicate terms need to be Subject, Process, Complement, and Adjunct, gives a clear picture of the system.

Participles and infinitives, as non-finite verb forms, do not typically function as Predicators within primary clauses; instead, participial and infinitival clauses function at the group level, a phenomenon known as “rankshifting,” or “embedding,” but they can fill any of the other three clause component roles.[291] This distributional difference forces the conclusion that the choice at the root of the system of ideational theme is the choice between placing a finite verb at the beginning of the sentence versus putting anything else there. Thus, the terms of the least delicate subsystem of ideational theme, which I will label subsystem 1, are +process and -process. Choosing +process places a verb at the beginning of the clause; choosing -process opens up subsystem 2. Subsystem 2 is the choice between +circumstance and -circumstance. The rationale for placing this choice here is that only substantives function as Subjects and Complements, whereas just about

[290] Porter and O’Donnell, “Greek Verbal Network,” 40.

[291] Technically speaking, so-called independent participles, i.e. a participle in whose environment occurs no finite verb on which the participle could depend, as well as similar infinitive constructions, could theoretically function as a Predicator, but these cases are sufficiently rare that I will ignore them for my purposes (cf. Porter, *Idioms*, 184). Porter (*Idioms*, 185–86) gives three passages as potential examples of independent participles with an imperatival sense: Rom. 12:9–19, 1 Pet. 2:18, and 1 Pet. 3:1. Interestingly, the annotators of these passages for the OpenText.org project classified none of these participles as Predicators, though in fairness they did so classify the two infinitive clauses in Rom. 12:15, Rom.c12_55 and Rom.c12_57.

anything can function as an Adjunct—including, occasionally, a substantive.[292] Choosing -circumstance opens up subsystem 3, which contrasts +/-actor. Choosing +actor places a Subject, typically realized as a noun in the nominative case, at the beginning of the clause; choosing -actor opens up subsystem 4, which contrasts +/-affected, distinguishing constructions where the action designated by the Predicator directly affects the participant realized by the Complement that is standing in thematic position (+affected), what has typically been called a direct object and that is typically realized by an accusative substantive in Hellenistic Greek, from two other constructions the OpenText.org model also classifies as Complements. Subsystem 5 contrasts the semantic choice behind these two constructions as representing a paradigmatic distinction: a "predicate nominative" realizes the choice +stative, and dative substantives realize the choice -stative as what have traditionally been called indirect objects.

The Area: Defining the Corpus

The two previous sections described the theoretical background for my project: SFL theory (in particular, register theory) stipulates that texts from similar genres exhibit similar patterns of grammatical choices, and the notion of grammatical probability provides a mechanism for quantifying these patterns.

[292] Ideally, I would break the category of circumstances down further, perhaps according to the question the particular Adjunct is answering, along the lines traditionally used for describing kinds of adverbial clauses—causal, locative, temporal, etc. (cf. Porter, *Idioms*, 230–43). Pursuing this further here, however, would mean giving up the main advantage of the network: allowing for some inevitable fuzziness around the edges, I can justify the rest of the hierarchy of choices—both those discussed already and those yet to be discussed—on differences in the distribution of particular forms in the language. By contrast, I can find no discernible correlation between particular structures and more delicate categories of circumstantial elements: multiple structures realize each of the traditional categories, and the same structure can realize multiple of the traditional categories (cf. Porter, *Idioms*, 230). Issues of lexis are at least partially to blame here. A unified lexicogrammar is "the grammarian's dream" (Halliday, "Categories," 54), but it is not yet the grammarian's reality. In my view, admitting the limits of a quantitative approach by not pushing the delicacy of the system any further is preferable to potentially compromising the validity of my statistical analysis by pushing further.

If, as argued above, studying the frequencies in text of forms that realize choices within linguistic systems can allow researchers to make illuminating statements regarding the question of the gospel genre, then the question of which texts to examine naturally arises. The answer to this question takes on even greater significance in view of SFL's commitment to explicitness and, consequently, its reliance on instances of natural language: arming oneself with such a tool and then neglecting to take equivalent care in selecting the data on which the tool is to be used would fritter away the very advantage that caused me to select it in the first place. Therefore, I will use the resources modern linguistics provides to ensure that the methodological rigor of the data collection process matches that of the model.

SFL matches corpus linguistics far better than many other potential options for a qualitative framework. First, SFL explicitly prioritizes authentic language use as a source for its descriptions.[293] Second, SFL's notion of instantiation explains the relevance of corpus data: textual data extracted from a corpus is relevant for grammarians because the grammatical system consists of the universe of potential texts.[294]

The sub-discipline within linguistics devoted to principles for distilling reliable generalizations from large amounts of natural language data is called "corpus linguistics."[295] In this context the term corpus refers to large groups of texts presumed to be representative of a particular language variety or varieties.[296] If the texts are in fact representative of the variety in question, then principles borrowed from statistics will

[293] E.g. He and Yang, *Absolute Clauses*, 54.
[294] Cf. He and Yang, *Absolute Clauses*, 55.
[295] Desagulier, *Corpus*, 7; cf. Gray and Biber, "Corpus," 138–52.
[296] Cf., e.g., Cantos, "Use of Linguistic Corpora," 99.

produce mathematically valid generalizations from data found within the corpus.[297] Corpus linguistic “methods have established themselves as among the most powerful and versatile tools to study language.”[298]

The proper design for a corpus varies depending on the research question the corpus is supposed to answer.[299] Since my research aims to determine whether the Gospel of Mark better matches referential or non-referential narratives written in Hellenistic Greek, the language varieties relevant to my study are referential narratives in Hellenistic Greek, non-referential narratives in Hellenistic Greek, and Hellenistic Greek as a whole.

Although corpus linguistics did not achieve formal recognition as a category within linguistic study until the early 1980’s, ideas characteristic of what is now known as corpus linguistics have a much longer history.[300] According to the editors of a recent conspectus on corpus linguistics, one can trace its roots to the creation of biblical concordances in medieval times.[301] Other influences, which they consider of even greater significance, include lexicography and structural linguistics as it existed before Chomsky.[302] One can easily see a genealogical relationship between the citation slips upon which generations of lexicographers have relied and the KWIC (“key word in context”) lists that form the output of many modern corpus investigations, and American structuralists like Zellig Harris “were the forerunners of corpora” both in terms of

[297] Cf. Hernández-Campoy and Schilling, “Application of the Quantitative Paradigm,” 63–75.
[298] Gries and Newman, “Creating and Using,” 257.
[299] Reppen, “Building,” 31.
[300] McCarthy and O’Keeffe, “Historical Perspective,” 5; cf. O’Donnell, *Corpus*, 38–76.
[301] McCarthy and O’Keeffe, “Historical Perspective,” 3.
[302] McCarthy and O’Keeffe, “Historical Perspective,” 4; cf. Tognini Bonelli, “Theoretical Overview,” 14–15.

procedure and "commitment to putting real language data at the core of what linguists study."[303]

Perhaps the most significant influence on modern corpus linguistics has been technological in nature: the introduction of the computer marked a turning-point in the history of linguistics, as was the case in many disciplines.[304] Advances in computer technology—more specifically, the technology for digitizing texts (i.e. the raw material of which a corpus consists)—mark the divisions between the three periods into which Tognini Bonelli divides the history of corpus linguistics: the first period in which texts had to be manually typed gives way to the second stage when the advent of the scanner and computer typesetting allowed for quicker access to texts, and around the turn of the millennium a third stage came about where corpus linguistics began using texts like blogs, emails, and sites from the Internet, which had never had an existence in physical form.[305] Since decreasing the amount of time and energy researchers must expend on collecting their data expands the amount of data they can accumulate within a reasonable timeframe, the size of corpora has ballooned commensurately with these labor-saving technological advances, representing perhaps the most obvious impact of technology on corpus linguistics.[306] Original, so-called "first generation" corpora from the early 1960s, such as the Brown corpus, consisted of one million words, whereas now corpora can consist of over a billion words of text even without considering the practically limitless amount of language data available by treating the Web as a corpus.[307] Tognini Bonelli

---

[303] McCarthy and O'Keeffe, "Historical Perspective," 4.

[304] McCarthy and O'Keeffe, "Historical Perspective," 6; Tognini Bonelli, "Theoretical Overview," 15.

[305] Tognini Bonelli, "Theoretical Overview," 17.

[306] Cf. McCarthy and O'Keeffe, "Historical Perspective," 6.

[307] McCarthy and O'Keeffe, "Historical Perspective," 6.

cites the following as a maxim of John Sinclair, one of the pioneers of corpus linguistics: "by about 1990 linguistics had changed from a subject that was constrained by a scarcity of data to one that was confused by more data than the methodologies could cope with."[308]

Corpus linguistics has much to offer, but it is not a panacea to cure all linguistic ills. Convincing descriptions of grammar will not emerge from corpus data alone, although some grammarians, styling themselves as "corpus-driven" in contrast to "corpus-based," have attempted to produce such entirely quantitative grammars.[309] In my view, these corpus-driven grammars fall into the same trap that waylays attempts to answer linguistic questions without invoking a linguistic model.[310] Ignoring questions of theory and methodology does not mean that one has successfully avoided them but that one has simply adopted ad hoc ones without reflecting on them or—just as importantly—informing the reader of what the answers are and the reasoning that led to them. For example, Elena Tognini Bonelli, one of the foremost proponents of corpus-driven grammar, implicates a theoretical category, namely instantiation, in her description of the purpose of corpus linguistics: "the aim of corpus linguistics can be seen as the analysis and description of *language use, as realised in text(s)*."[311] As I have stated repeatedly with regard to previous quantitative attempts to study the question of the gospel genre, quantitative data without a qualitative framework surrounding it is often of unclear significance.

---

[308] Tognini Bonelli, "Theoretical Overview," 16.

[309] According to Desagulier (*Corpus*, 9), "[c]orpus-driven linguistics is the radical extension of corpus-based linguistics: the corpus is not part of a method but our sole access to language competence."

[310] Cf. Desagulier, *Corpus*, 7–8.

[311] Tognini Bonelli, *Corpus*, 2.

Returning to the mapping metaphor, one determines what features a map will show and, correspondingly, the amount of work necessary to generate it on the basis of what the map is intended to illustrate. Expending the effort to incorporate extraneous details into a map is a waste, but not expending the effort to incorporate details necessary for a map's intended purpose would result in a useless map. For instance, contrast the sort of map one might use to guide a friend to a particular place in town with the sort of map an orienteer might use to find his or her way across trackless backcountry: in the former case, the map probably consists of some lines representing roads and maybe some landmarks, whereas maps of the latter type encapsulate much more information but at the same time are far more complex both to generate and to read. Orienteering maps vary among themselves in terms of scale: one can choose a high-scale map that shows lots of detail or a low-scale map that shows lots of territory. Theoretically, one could have a map with both detail and expanse, but the resulting size would render it unusable.

The relevance of this discussion of maps is that building a linguistic corpus deals with some of the same issues as generating a map. Corpora of naturally-occurring language data exist in order to represent a particular variety (or varieties) of language. The samples of which a corpus consists of a "subgroup of people that reflects the population as a whole (in terms of their social and linguistic characteristics), and therefore lends itself to generalizations above and beyond the scope of the study."[312] Corpus designers also face a trade-off between extreme detail and practicality.[313] Ideally, the corpus would consist of every instance of the target population, rendering the question of representativeness moot, but any conceivably interesting linguistic population

[312] Buchstaller and Khattab, "Population," 74; cf. Schneider, "Historical Variation," 68.
[313] Buchstaller and Khattab, "Population," 75.

will be too large to make investigating it this way practical.[314] The situation in this regard is even more serious for the present study because such a corpus of Hellenistic Greek is not even theoretically possible, given that the majority of the texts written in it have perished (to say nothing of the fact that modern researchers have little, if any, access to its spoken form).[315] Some pruning of the data is needed, but the pruning cannot be haphazard if the results of analyzing the corpus are to remain convincing.[316]

A corpus designer tries to represent accurately the language behavior of a target population, just as a cartographer is trying to represent accurately a given piece of terrain. A cartographer produces a functioning map by knowing what sort of information the map is expected to display and then surveying the piece of terrain to be represented in order to obtain the necessary data. Corpus designers likewise have to decide on criteria that they think will produce samples representative of the target population and then apply those criteria to generate the corpus.[317] Neither map nor corpus can exist without both a qualitative framework and quantitative implementation.[318]

Two facets of the quantitative aspect of corpus design demand attention here: (1) ensuring that the corpus samples enough texts to be representative of the target population and (2) ensuring that each sample is sufficiently long to be representative of

---

[314] Buchstaller and Khattab, "Population," 74; O'Donnell, *Corpus*, 108–12.

[315] Halliday ("Spoken," 5) refers to speech as the "mainspring of semogenesis [i.e. meaning making—NB]" because it expands "the frontiers of meaning potential." Furthermore, Hernández–Campoy and Schilling ("Application of the Quantitative Paradigm," 68) note that "rely[ing] on written sources constrains the probability of variation." Modern corpus researchers of Hellenistic Greek should take note of the role speech plays in language change. Hellenistic Greek no longer has any living speakers, but it was constantly changing when it did. This implies that a corpus-based investigation of Hellenistic Greek should define its population as narrowly as possible in order to avoid introducing extraneous variables.

[316] Every corpus is, in statistical terms, a convenience sample ("Population," 83–84), but "the theoretical claim of most research in this field would be that large numbers of observations allow for making generalizations in a statistical sense" (84).

[317] The examples Schneider ("Historical Variation," 59–60) offers are particularly apropos in this regard.

[318] Cantos, "Use of Linguistic Corpora," 100.

the text from which it originated.[319] In both cases, "the foundational task" is accumulating enough observations to produce mathematically reliable results.[320] I will illustrate both quantitative facets (in reverse order), with O'Donnell's preliminary corpus for studying Hellenistic Greek.[321]

O'Donnell prefers to use entire works as corpus samples (rendering the question of sample representativeness moot), but in cases where a work he wishes to include is long enough to dwarf the New Testament documents he samples twenty thousand words, chosen such that they constitute an identifiable section of the work in question.[322]

The question of how many samples it takes to create a representative corpus is more complicated, since it depends on how broad the target population is. Homogenous populations exhibit less variance, so fewer samples will be required to represent them accurately; broader populations will need more samples so that all of the variety will be represented in the corpus.[323] For example, O'Donnell's corpus aims to represent Hellenistic Greek in its entirety, so he has to include a large number of samples in order to represent the broad swath of Hellenistic Greek, both in terms of language formality—from vulgar papyri to Atticizing prose—and in terms of genres.

One will never be able to demonstrate the representativeness of a historical corpus in the pure statistical sense; too many of the needed criteria are simply unknowable.[324] In

---

[319] Desagulier (*Corpus*, 3) describes this aspect as corpora being "a representative and balanced sample of representative and balanced samples."

[320] O'Donnell, *Corpus*, 102. Woods and his co-authors lay out the relevant mathematical background (*Statistics*, 37).

[321] O'Donnell, *Corpus*, 132–37. He provides a tabular list, which is easier to reference, as an appendix (164–65); cf. Gray and Biber, "Corpus Approaches," 141.

[322] O'Donnell, *Corpus*, 114.

[323] Cf. Desagulier, *Corpus*, 4.

[324] Cf. Hernández–Campoy and Schilling, "Application of the Quantitative Paradigm," 66; Schneider, "Historical Variation," 76.

this respect, the historical corpus is no different from history in general: we can claim to have reliable historical knowledge about some things without having to claim we have exhaustive historical knowledge and, by the same token, we can claim to have reliable knowledge of a historical stage of a language without having to claim that we have identified and controlled for every possible form of variation.[325] Even in cases where the population is directly observable, "achieving genuine statistical representativeness is extremely difficult."[326] Fortunately, doing so "is not always necessary to obtain solid results and revealing insights."[327]

O'Donnell's corpus design accounts for as many of the variables as possible and is thus a worthy beginning point for further corpus-based study of Hellenistic Greek with regard to both the qualitative and quantitative aspects of its design. In an ideal world, I would simply adopt it, rather than developing my own.[328] However, the familiar trade-off between comprehensiveness and practicality again raises its ugly head here: at 512,301 words, O'Donnell's corpus is too extensive for manually examining it to be practical in a realistic timeframe, especially since many of the texts would have to be annotated before analysis itself could take place.[329] The qualitative framework of O'Donnell's corpus can remain, but the size had to be cut down drastically. I proceeded along three avenues for doing this without harming the representativeness of the corpus: (1) I excluded non-

325 Hernández–Campoy and Schilling, "Application of the Quantitative Paradigm," 70; cf. Cantos, "Use of Linguistic Corpora," 104–5.

326 Hernández–Campoy and Schilling, "Application of the Quantitative Paradigm," 65; cf. Desagulier, *Corpus*, 5.

327 Hernández–Campoy and Schilling, "Application of the Quantitative Paradigm," 65

328 Cf. Reppen, "Building," 31.

329 The time-consuming nature of manual analysis and the difficulty of writing automatic parsers that can reliably find all the instances of a particular system are the two main factors holding back the application of corpus linguistics to studies of grammar (cf., e.g., Halliday, "System and Instance," 79–80; Halliday and James, "Quantitative Study," 95).

narrative texts, since Mark is clearly a narrative, (2) I restricted the corpus to texts of likely Jewish provenance, and (3) I decreased the size of each sample. Whereas O'Donnell aimed to represent Hellenistic Greek in its entirety, this more restricted corpus only aims to represent Jewish narrative texts in Hellenistic Greek; a more limited corpus can accomplish this restricted goal. I now turn to defending in detail each of these delimitations.

O'Donnell's corpus contains texts from a wide variety of genres, including quite a few non-narrative ones that clearly do not relate to the Gospel of Mark, since—whatever else it might be—Mark is clearly a narrative.[330] O'Donnell needed these texts to accomplish his goal of describing Hellenistic Greek as a whole, but the best that can be said for incorporating them into my study is that contrasting their lexicogrammatical patterns with those of Mark might allow me to infer what parts of Mark's patterns come simply from its status as a narrative (rather than some specific type thereof). However, that is not the primary question in which I am interested, and following that tangent would increase the workload dramatically without measurably affecting the results for my primary inquiry.[331]

Limiting the corpus to works of likely Jewish provenance also simplifies the study.[332] Although a few scholars have argued that the Evangelist was a Gentile, Collins correctly calls attention to both the tenuous logic behind those arguments and the value of

[330] Mark's status as a history—to say nothing of its historicity—does not follow from its narrative status: histories tend to take narrative form, but it does not follow that all narratives are histories; thus the inclusion of other narrative genres in the corpus (cf. Güttgemanns, *Candid Questions*, 21–22).

[331] Cf. Eddington, *Statistics*, 9.

[332] Contrast Crellin's (*Syntax and Semantics*, 20) choice of texts, which is dominated by Gentiles: Polybius, Josephus, Plutarch, Appian, Meander, Dionysius Thrax, Philodemus, Diodorus Siculus, Dionysius of Halicarnassus, Strabo, Philo, Aristonicus of Alexandria, Babrius, Pausanius, Phrynicus, Aelius Herodinaus, Cassius Dio, and Galen.

the Aramaic loanwords sprinkled throughout the Gospel for ascertaining the author's background.[333] Specifically, familiarity with Aramaic indicates that the author hailed from the eastern part of the Empire and, I would add, was likely a Jew. If the author was a Jew, then Jewish texts would seem the most likely candidates for useful parallels.

Reducing the length of each sample without harming the corpus's representativeness is necessary in order for my project to be practical. Investigating clause-level systems should allow for shorter samples.[334] Most corpus-based work to date has investigated lexis or grammar easily accessed from lexis.[335] O'Donnell's work is no exception. Practicality necessitates that this be the case. Such a desire for plucking low-hanging fruit is understandable—particularly in the case of a relatively young discipline like corpus linguistics where even the bottom branches may not yet be entirely picked over—but in this case it has led to what Halliday terms the "lexicogrammatical bind."[336] Despite most agreeing that corpora should be just as pivotal in studying grammar as they have proven to be for investigating lexis, few have attempted to implement it, and those few have typically restricted themselves to function words that "tell us very little about what is going on underneath."[337] Halliday posits an inverse relationship between the importance of a pattern and the ease with which one can find it.[338] As with geometry, it turns out that there is no royal road to corpus grammar: researchers who want to obtain significant results must undertake the tedious work of annotating their corpus at levels of

---

333 Collins, *Mark*, 6.

334 Cf. Halliday, "Quantitative Studies," 131. Desagulier (*Corpus*, 6) puts the situation this way: "Small [corpus] size becomes a problem if the unit you are interested in is not well represented. All in all, size matters, but if it is wisely used, a small corpus is worth more than a big corpus that is used unwisely."

335 Halliday, "Spoken," 17–18; Halliday, "Working," 43.

336 Halliday, "Spoken," 18; cf. Halliday, "Towards Probabilistic," 59.

337 Halliday, "Spoken," 18.

338 Halliday, "Spoken," 18; Halliday, "Corpus Studies," 67.

discourse above the word, and for the foreseeable future this work will have to be done manually.[339]

The third delimitation is more complicated than the other two because the mathematical principles governing representativeness of samples specify the number of observations needed, rather than a total number of words.[340] A mathematically reliable sample consists of 385 units.[341] In the case of my project, the units in question are primary clauses, because unembedded secondary clauses are sufficiently rare that using samples long enough to contain a statistically significant number of them is impractical, and I exclude embedded clauses because they have a different set of thematization options than other sorts of clauses. The number of words needed to contain a statistically significant number of clauses is 385 multiplied by the number of words between usable tokens. Determining the number of words between usable tokens is complicated, because clauses do not consist of a fixed number of words and primary clauses occur in no fixed distribution relative to the other two types of clauses. Given these limitations, the most reasonable course seems to be working backwards, choosing a sample length and seeing if it allows for a reasonable expanse of text between predicators of a primary clause. Samples of ten thousand words would have the virtue of being about as long as Mark and of economizing on O'Donnell's corpus by at least 50%. Adopting this sample length allows for approximately twenty-five words between usable tokens, which seems more than reasonable as an approximation for the maximum distance between the predicators of two primary clauses.

---

[339] Halliday, "Corpus Studies," 67; cf. Jenset and McGillivray, *Quantitative Historical Linguistics*, 16.

[340] Schneider, "Historical Variation," 60.

[341] This figure comes from the formula Woods et al. provide (*Statistics*, 109).

"Corpus-based approaches to linguistic analysis are ideally suited to comparisons of the use of linguistic constructs across registers."[342] Corpus linguistics is "inherently a distributional discipline because" they answer three questions: (1) how commonly does something occur? (2) how close does something occur to something else, or (3) is there any pattern between occurrences in text and their extralinguistic contexts.[343] Corpus-based approaches have yet to catch on in historical linguistics to the degree they have in other branches of the discipline, but their use is on the upswing, which a number of voices welcome.[344] The value of corpora for studying genre is widely accepted.[345]

---

342 Gray and Biber, "Corpus Approaches," 142.
343 Gries and Newman, "Creating and Using," 274.
344 Jenset and McGillivray, *Quantitative Historical Linguistics*, 8.
345 Cantos, "Use of Linguistic Corpora," 101; Nevalainen and Raumoulin-Brunberg, "Historical Sociolinguistics," 24.

CHAPTER 3: SURVEYING TEXTS (PROCEDURE)

Chapter 1, surveying previous investigation into the Gospel genre, concluded that most recent investigations on the topic conclude that the literary milieu surrounding the production of the Gospels offers some useful analogies for them, but they disagree on what particular analogy is most appropriate. This disagreement seemed to stem from a failure to elucidate clearly the criteria for such a determination. Qualitative, impressionistic methods offer no hard data whose validity is intersubjectively valid; previous quantitative attempts, by contrast, have plenty of hard data but leave unresolved exactly why they relate to genre.

Chapter 2 covered my proposal for using systemic-functional linguistics (SFL) to fill the theoretical lacuna the literature review observed. SFL shores up previous quantitative attempts to describe the gospel genre in at least three ways. First, the notion of stratum provides a rigorous, clearly traceable link between concrete forms in a text and extratextual features that influenced them, such as genre. Similar situational constraints produce similar choices to express meanings appropriate for that situation and, thus, texts with consistent, identifiable features. Second, Halliday puts forward the notion of grammatical probability as a means of quantifying these consistently shared features: particular options in a grammatical system may be more appropriate than others for a given type of extratextual situation (a classic case being the imperative mood in recipes) meaning that the form realizing this favored option will occur more frequently in texts emerging from that situation than it does in the language as whole. Thus, if one can identify the systems whose relative frequencies, what Halliday calls "grammatical probabilities," are reset by a particular set of situational constraints, what Halliday calls a "register," then one has a quantifiable fingerprint by which to identify what texts belong to a certain register. Lastly, and in some ways the combination of the previous two, SFL puts forward the notion of "instantiation" as a means of putting theoretical rigor behind the notion of describing registers or, stated another way, genres. Register is only one of many factors that affect an author's choices in producing a text, so one text sheds only indirect light on the characteristics of its register. Conversely, if one could somehow count the frequencies of the choices from a particular system in every text belonging to a register, the proportion with which each choice was selected would, by definition, be equal to the grammatical probabilities associated with the system in that register. Instantiation refers to the level one is trying to describe on the scale from one text to all texts. In summary, SFL strengthens the methodology behind quantitative study of the gospel genre by explaining how concrete forms in a text relate to extratextual features like genre, by clearly defining the process of inductive reasoning by which a researcher moves from the properties of an individual text to the common properties that unite them as a group, and by providing a tool for moving from these theoretical principles to comparisons of groups of texts in a principled fashion.

This chapter describes the procedure for applying the theoretical framework described above. One linguist describes this sort of study: "Corpus-based variation studies focus on the comparison of the use of objectively countable linguistic features. That is, variationist studies tend to concentrate on the quantitative analysis of extracted tokens. Typical variationist research questions take the form: is variable x used differently in corpus A compared to corpus B."[1] The basic logic behind this process is the

---

[1] Cantos, "Use of Linguistic Corpora," 103.

same as other quantitative projects in linguistic research. According to the introduction of a recent journal issue dedicated to improving the practice of such research in applied linguistics, this research paradigm assumes that by setting out the standards by which they "collect, analyze, and interpret numeric data, researchers should be able to estimate with considerable accuracy the magnitudes and other patterns in language phenomena that interest them" and, by the same token, they "should be able to match the complexity of language phenomena with" appropriate methods for making sense of these phenomena through inferential statistics.[2]

The Area to Map: A Corpus of Prepared Jewish Texts

This section lays out the components of the corpus to which I compared Mark. It is divided into three subsections. The first provides an overview of the corpus as a whole. Each of the texts in the corpus is then discussed in either the second or third subsection, depending on the sub-corpus to which it belongs. The discussion of each constituent text explains how it meets the design objectives as well as the edition of each text I used.

Overview of Corpus Design

The first procedural question is: what texts am I mapping? Halliday reasonably cautions that "a corpusdriven [sic] grammar needs a grammar-driven corpus."[3] Presumably, the same would be true of a corpus-driven semantics, so the corpus must be designed with the hypothesis I wish to test in mind. I want to determine which candidate analogical genre is most similar to Mark, so the corpus must focus on the most promising analogies discussed above, namely history, biography, and novelistic works.

---

[2] Norris et al., "Improving," 3.
[3] Halliday, "Spoken," 20.

Differentiating between these three genres is often difficult, however.[4] Stadter notes: "we can only speak of separate genres of history and biography if we remain aware of the fluidity of the boundary between them."[5] The border between history and novel is also somewhat unclear. Statistical analyses require discrete categories, so I need to make some sort of clear distinction, even if they are purely notional.

I think Gregory E. Sterling's continuum of Jewish narrative offers a solid way forward. He divides extant Jewish narratives into three categories: histories, historical novels, and "prose fictions."[6] Histories are explicitly based on sources and the other two categories are "free creations—even if they use known figures."[7]

Thus, my corpus consists of two parts, representing the extreme ends of Sterling's continuum. The first part combines both history and biography to constitute "referential narrative."[8] The second part, which I refer to as "non-referential narrative," consists of samples derived from Sterling's prose fiction category. This is similar to the approach advocated by Edgar Schneider for studying historical variation in other kinds of texts.[9]

Sub-Corpus 1: Referential Narrative

As discussed above, the first half of my comparative corpus consists of what I am calling "referential narrative." Five of the histories Sterling listed are potentially long enough to provide a useful sample: 1 Esdras, 1 Maccabees, 2 Maccabees, *Antiquities of the Jews*, and Philo's *Embassy to Gaius*.[10] Additionally, Sterling lists three biographies that could

[4] Cf., e.g., Stadter, "Biography," 2:528–2:529, 2:540; Morgan, "Fiction and History," 2:563–564.
[5] Stadter, "Biography," 2:528.
[6] Sterling, "Jewish Appropriation," 1:232.
[7] Sterling, "Jewish Appropriation," 1:232.
[8] Cf. Becker, "Mark," 127.
[9] Schneider, "Historical Variation," 60–61.
[10] Cf. Sterling, "Jewish Appropriation," 1:233.

round out the referential samples: Philo's *On Joseph*, the autobiography of Josephus (*Life*), and *The Lives of the Prophets*.

Of these, I focused on 1 Esdras and 1 Maccabees. I decided that the composite nature of *The Lives of the Prophets* made investigation of clause thematization problematic: none of the sections were sufficiently long in and of themselves, so I would have had to use two or more, raising the question of how the sections relate to one another. I did not find a convenient electronic edition of the works of Josephus or Philo to serve as a point of departure for the mapping process described in the next section.

The book of 1 Maccabees recounts how the Jews attained independence from the Seleucid empire during the middle of the second century BCE.[11] In terms of structure, Bartlett divides the book into three sections: an introduction (1:1—2:70), narratives focused on Judas Maccabeus (3:1—9:22), and narratives focused on his successors Jonathan and Simon (9:23—16:23).[12] He subdivides the first of these major sections into two parts, namely 3:10—6:63 and 7:1—9:18, although he allows that chapters 5 and 8 "clearly interrupt the sequence of the surrounding narrative."[13] Given that chapter 5 interrupts the narrative and chapters 1–4 provide enough data—at least for the narrative framework—to meet the size of a conservatively large sample, I chose chapters 1–4 as my sample.[14] I used the Logos morphologically-tagged edition of Rahlfs' *Septuaginta* as the base text for this sample.

---

[11] Cf. Bartlett, "1 Maccabees," 807–30.
[12] Bartlett, "1 Maccabees," 807.
[13] Bartlett, "1 Maccabees," 811.
[14] As it turned out, this choice proved less than optimal. See "Lessons Learned" in Chapter 8.

The book of 1 Esdras condenses roughly a century and a half of Jewish history, from Josiah's reforms to the return from exile led by Ezra, into nine chapters.[15] According to Michael F. Bird, "the literary form of 1 Esdras is strictly speaking a historical narrative," although he classifies it more specifically "as 'Rewritten Bible.'"[16] While the extant form of 1 Esdras is generally considered to be a translation of a Semitic original, according to Bird, it "is written in good quality Greek," as opposed to the "mechanical and wooden translation" found in 2 Esdras.[17] Thus, unlike a slavishly literal translation where Hebrew clause structure might interfere, 1 Esdras should provide reliable data on the Greek thematization system. The sample I used consisted of chapters 1, 2, and 5:4—9:55. In other words, it consists of the whole book, except for the story of Darius and his bodyguards.[18] This sampling choice makes sense because, as scholars have long noted, the story of Darius and his bodyguards is "unique to 1 Esdras and stands apart from the rest of the narrative."[19] I used the Logos morphologically-tagged edition of Rahlfs' *Septuaginta* as the base text for this sample.

Sub-Corpus 2: Non-Referential Narrative

As discussed above, the second half of my comparative corpus consists of what I am calling "non-referential narrative." Sterling lists a number of works in his prose fiction category.[20] Of these, *1 Enoch* is of doubtful relevance because it is most likely a translated text, and one that is only partially extant in Greek at that.[21] The *Testaments of*

---

[15] Bird (*1 Esdras*, 6) gives the dates of 621 BCE and 458 BCE for the earliest and latest events, respectively.

[16] Bird, *1 Esdras*, 8.

[17] Bird, *1 Esdras*, 22.

[18] Williamson, "1 Esdras," 853; cf. Bird, *1 Esdras*, 21–22.

[19] Bird, *1 Esdras*, 2; cf. Williamson, "1 Esdras," 853–55.

[20] Cf. Sterling, "Jewish Appropriation," 1:233.

[21] Cf., e.g., Isaac, "1 Enoch," 6.

the Twelve Patriarchs fall into the same pitfall I discussed above with reference to *Lives of the Prophets*. *Susanna* is too short to provide an adequate sample. The *Testament of Solomon* is likely to be too late.[22] Thus, the non-referential portion of the corpus consists of *Tobit*, *Judith*, and *Joseph and Aseneth*.

Tobit's genre is a matter of debate, but it is often considered an example of the "Jewish novel."[23] *Tobit* is the backbone of this portion of the corpus because I included the entire book, rather than choosing a sample. I used the Logos edition of Rahlf's *Septuaginta*.

Judith is also an important component of the corpus. Gerald West characterizes the book of Judith as "a powerful, compelling, and somewhat unsettling narrative."[24] This narrative, as West points out, divides evenly into two halves, which he calls "acts"—chapters 1–7 and chapters 8–16.[25] Despite the fact that Judith does not arrive on the scene until "the second act," West argues that "Act 1 is an integral part of the narrative."[26] This being the case, and since Act 1 is almost exactly the size of the conservatively large sample discussed in Chapter 2, I have chosen to use it as my sample. As with the other samples from the Septuagint, I used the Logos edition of Rahlf's *Septuaginta*.

Joseph and Aseneth forms the final component of the corpus. This novelistic work also falls into two parts, although in this case the parts are of uneven length: Section 1 consists of chapters 1–21 and Section 2 is chapters 22–29.[27] I chose Section 2 for my

---

[22] Cf. Duling, "Testament of Solomon," 940–43.
[23] Grabbe, "Tobit," 736.
[24] West, "Judith," 748.
[25] West, "Judith," 748.
[26] West, "Judith," 748.
[27] Burchard, "Joseph and Aseneth," 182.

sample. I used the Greek text available, with morphological tags, within Logos Bible Software.

Details of the Mapping Process

Having determined the texts to map, the next stage is to lay out the process of mapping itself. This process consists of two main parts. First, I must apply the projection to the area, so to speak. This means putting the samples into a form where I can reconstruct the choices that generated them. The second stage is cataloguing these choices to portray the system of which they are the output. The two subsections below describe both phases in more detail.

Applying the Projection to the Area: Corpus Annotation

As discussed in Chapter 2, SFL theory posits that texts result from a set of choices from describable, closed systems, and which choices a language user makes on a particular occasion depends on what he or she is trying to accomplish, hence a view of language that is both systemic and functional. From this perspective, genre variation manifests as transitions in grammatical probabilities, which are derivable from frequencies observed in texts. The frequencies I intend to test are thematization choices, so I need to break the comparative corpus into clauses so that I can determine what sort of clause constituent is at the beginning.

Clause breakdowns and other markup of plain data files "added to provide specifically linguistic information" fall under the heading of corpus annotation.[28] Corpus annotation invariably involves analysis. Given that my project is comparative, I need to

[28] Gries and Newman, "Creating and Using," 263; cf. Jenset and McGillivray, *Quantitative Historical Linguistics*, 10–12.

make sure all the analyses proceed from the same bases in order to make sure the comparisons are appropriate. This necessitates using the same set of annotation guidelines for the entire project. I used the guidelines developed for the OpenText.org project.[29] In turn, these guidelines assume a Hallidayan perspective on language, so aspects of SFL come into play as well.

One recent investigation of NT clause structure notes the suitability of OpenText.org for this sort of inquiry.[30] Furthermore, using the same guidelines as others who have previously annotated the Greek of the NT means that the box diagrams that I create for this project should be useful for other research questions, allowing me to contribute indirectly to those areas as well, in addition to the specific question with which I am concerned.[31]

The annotation process begins by separating the clauses within a text from each other, and doing this means determining what constitutes a clause. A prototypical clause consists of a verbal form and all the material related to it, so the first task in clause annotation is to find all the verb forms and determine what parts of the text relate to each of them. The material unaccounted for during this preliminary pass will largely fall into two categories, namely verb-less clauses and cases where the reader is expected to retrieve the verb from the previous clause.

The second step consists of determining the category of clause to which each clause belongs. There are three possibilities: primary clauses, secondary unembedded clauses, and secondary embedded clauses.[32] Primary clauses and secondary unembedded

[29] O'Donnell et al., eds., "Clause Level," July 7, 2004; cf. Tan, "Guide," February 3, 2006.
[30] Stevens, "Clause Structure," 65.
[31] Kendall, "Data," 43–45.
[32] Cf. Halliday, *Grammar*, 195.

clauses both typically have finite verbs as their predicator, but secondary unembedded clauses typically have a subordinating particle as well. Nonfinite verbs (participles or infinitives) typically indicate that a particular clause is a secondary embedded clause, although OpenText.org annotates some specific participial constructions (genitive absolutes) and infinitival constructions (articular infinitives combined with a preposition) as unembedded secondary clauses.[33] Ultimately, the distinction between unembedded and embedded secondary clauses rests on whether or not they are "rankshifted" down so that they play a discernible role in the structure of another ranking clause; the association between clause type and verbal mood is only a rule of thumb.[34]

The above-mentioned differentiation between types of secondary clauses on functional grounds calls attention to the next step of clausal annotation, namely determining how the parts of a clause relate to each other. Just as I differentiate between embedded and unembedded clauses on functional grounds, I also only break down clausal components to a level that can be shown to play a functional role in the clause.[35] OpenText.org uses a graphical tool called a box diagram to label how various parts of a clause are functioning. Below is OpenText.org's box diagram for the first part of Mark 1:45. After describing the various parts of this diagram, I will use it to illustrate the procedure I performed on primary clauses in the comparative corpus.[36]

---

[33] Tan, "Guide," February 3, 2006.

[34] Matthiessen et al., *Key Terms*, 170; cf. Halliday, "Machine Translation," 28–29.

[35] Halliday, *Grammar*, 22–24.

[36] As mentioned in the previous chapter, this study comprehensively treats only primary clauses. Unembedded secondary clauses lie completely outside my project, and embedded secondary clauses are only involved as components of primary clauses in which they may be embedded.

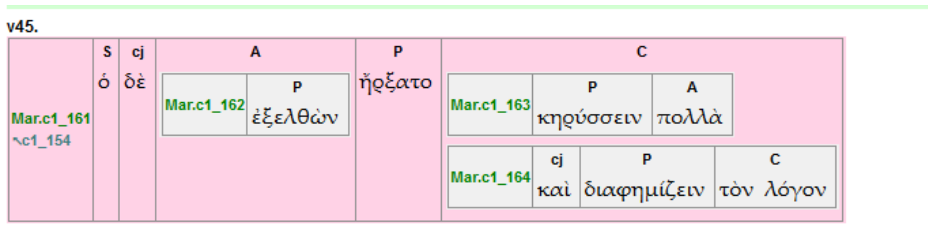


Figure 1: Example Box Diagram

OpenText.org stores separate chapters in separate files, so the label at the top, “v45,” clearly identifies this particular box diagram as coming from Mark 1:45. However, a verse often—perhaps even usually—contains more than one clause, so OpenText.org also provides a unique identifier for each clause, consisting of an abbreviation for the book (Mar), the chapter (.c1), and a number (161–164, in the case of the example). The clause identifier in small print with an arrow next to it references the last primary clause outside of projected discourse. The colors signify the type of clause: pink means primary clauses, and the bluish green signifies embedded secondary clause. The vertical lines set off the various clause constituents, typically either a word group or an embedded clause rankshifted down to word-group status, each of which will have a particular function in the structure of the clause. Even from this specific example one can see that the number of words of which a word group consists varies widely: most of the word groups within Mar.c1_161 consist of one word, while the final word group includes a total of six words spread across two secondary embedded clauses.[37]

Only one feature of the example diagram is left for me to describe, and it is the central one for my purposes. The letters above each word group label the role that word group plays in the structure of the clause.[38] Mar.c1_161 illustrates these labels well

---

[37] Cf. Halliday, *Grammar*, 25–26.
[38] Halliday, *Grammar*, 26–27.

because five of the six possible labels occur here. The clause begins with a Subject (S). An author includes a Subject constituent when he or she wants to provide more information about the person(s) or thing(s) engaged in the action of which the clause speaks than the verb's ending would supply. Then an inter-clausal conjunction (cj) occurs. The third word group is an Adjunct (A). Adjuncts specify "circumstances associated with the process," such as location, time, or manner.[39] Fourth comes the Predicator (P), which details the process going on in the clause (in this case, "beginning."). The clause winds up with a Complement (C). Complements specify who or what the process affects (i.e. they are primarily what traditional grammarians have called direct and indirect objects.) The sixth possible role, absent from this clause, is direct address (add).

Having explained the various parts of an OpenText.org box diagram, the text task is to explain how to go about producing one for a text that has not yet been annotated. Before labeling one will have to break up the clause into clause constituents. Once one turns to labeling, one should begin with the Predicator, since Predicators correspond to verbal forms, and the presence of a verbal form is likely what signalled the presence of a clause in the first place. Next, one looks for a noun in the nominative case. If one is present, the group of which it is a part will likely be the Subject. Similarly, nouns in the accusative or dative case likely signal a Complement, although—as the example diagram shows—other structures like infinitives can serve as the Complement of certain verbs. A variety of forms can signal the presence of an Adjunct.

[39] Tan, "Guide," February 3, 2006.

These morphological rules of thumb normally make categorizing the ideational theme of a particular clause relatively straightforward. They are not 100 percent effective, however, and some instances could be analyzed in multiple ways. In such cases, I used the OpenText.org diagrams produced for the NT as a guideline so that my annotation would be as compatible with those diagrams as possible.[40]

Drawing the Maps

With the qualitative framework in place, the stage is set for beginning to collect quantitative data, specifically the frequency counts of forms that realize the options within the systems in which I am interested.

This annotation process will produce a set of box diagrams like the ones OpenText.org provides for the New Testament for all of the texts in the corpus. However, since OpenText.org already provides box diagrams for Mark, I will use those rather than reinventing the wheel.

These box diagrams are the framework for generating grammatical probabilities, which is the second major step of my procedure. As explained above, viable probabilistic modeling of grammar relies on choices the language forces the author to make in the communicative process, where every clause is forced to select an option from a system network that specifies the set of potential meanings in that particular system. Every form occurring within the domain of the system network should result from a path through the system network and only one path through the system network should produce that form. Under such conditions, one can justifiably reason backwards from the forms observed in text back to the choices that generated them and, given a large enough sample of text, the

[40] Cf. Kendall, "Data," 46–47.

relative frequencies of the forms realizing a choice in the system will approximate very closely the grammatical probability of the choice in question.

Having laid out such a system network in the previous chapter, the task at hand is generating the relative frequencies that will serve as the basis of my approximations. The first step in determining a relative frequency of a form is determining its absolute frequency, i.e. the number of times it appears. The second step is determining the number of times it could have appeared. With these two pieces of information in hand, computing the relative frequency is as simple as dividing the absolute frequency by the number of times it could have appeared. I now turn to describing both preliminary steps in more detail.

For my purposes, the absolute frequencies in which I am interested are the number of times each kind of clause constituent described above (Predicator, Adjunct, Subject, Complement) appears as the first constituent of a primary clause. Each primary clause selects a first component from this closed system, so the logical way to count how many times a particular type of clause constituent serves as ideational theme would be to look at the first primary clause identified in the box diagrams of a text sample, record which component type the first clause component is, do the same for the second primary clause, and repeat this process until one reaches the final primary clause identified in the box diagrams of a text sample. The process then repeats for all the samples in the project. This process is theoretically defensible, since it mirrors the relationship between paradigmatic systems and syntagmatic order in SFL theory: each point at which a choice is available in the unfolding of syntagmatic order offers an opportunity for any of the paradigmatic choices to appear. I have already explained the box diagram component, so now I turn to describing the process for recording the choices found within them.

I recorded the thematization choices encountered in the process of proceeding through the box diagrams in a Microsoft Excel spreadsheet.[41] The design of this spreadsheet evolved over the course of the project as I accumulated experience and solved problems I was having. The original design of the spreadsheet had the data for Mark and all the comparative samples laid out in a single file. The rows represented individual texts, and each choice had its own column. Recording a choice was as simple as increasing the number in the cell at the intersection of row and column by 1: if, for example, I found a thematic Predicator in Mark, the number in the cell at the intersection of Mark and Predicator increased by 1. I realized very quickly, however, that keeping all the data in one file was not going to work because chunks of text long enough to generate statistically reliable samples exceed the length I can count without making a mistake or, at least, wondering if I had accidentally skipped a clause or counted one twice. As a result, I decided to keep one master file where the final data went and a separate set of files, one for each text, that contained the data for that text. Each of this latter group of spreadsheets broke down the data from the associated text on a chapter-by-chapter basis, along with a final tab arranged like the previous ones but whose cells aggregate the frequencies observed in the individual chapters.

This procedural innovation had several beneficial effects. First, having the data for each text in a separate file, rather than rows in one file, meant that I no longer had any use for the row dimension, allowing me to collect the data on the narrative framework

[41] Cf. Smith and Seoane, "Categorizing," 218–21; Gries, "Elementary Statistical Testing," 363–66.

and projected discourse separately. I wanted to do this anyway but had no clear way to implement it until I freed up the row dimension. As it turns out, this improvement was extremely fortuitous for the outcome of this project. [42] Secondly, having the data available in smaller chunks of one chapter made verifying my results easier: before moving to the next chapter, I would verify the total number of clauses against the clause labels in the box diagrams to make sure I had neither missed a clause nor counted one twice and then spot-check one or two of the column counts, figuring a spot-check of individual counts would be sufficient to turn up problems, if the total was accurate.

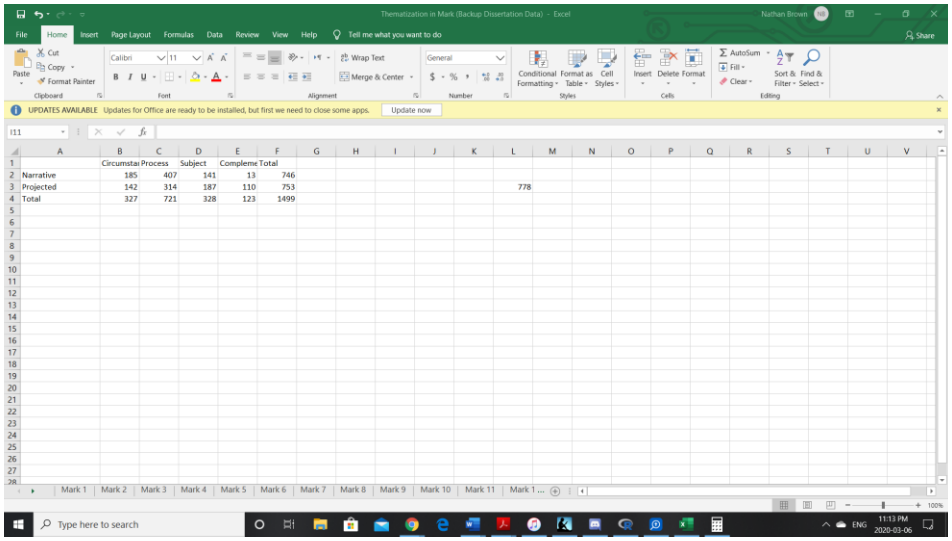
An example of the product resulting from pursuing this procedure through an entire text sample follows below (Figure 2). Notice that the filename at the top of the screenshot clearly identifies the sample under consideration and identifies this as the backup copy. This is to differentiate it from the master copy kept together with the data from other samples, which serve as the basis of the tables in chapters 4 through 6. As stated above, each tab contains the frequency count for a particular chapter (e.g. "Mark 1" in the list of tabs towards the bottom of the screenshot) and a single tab that aggregates the individual counts, the latter of which is shown below. Unlike the individual chapter frequencies, which I incremented manually as I went through, an automated summation formula calculates these totals.[43] The rows cross-classify the frequencies between the narrative framework ("Narrative") and projected discourse ("Projected), and the columns are the four main types of clause constituents. Chapters 4 through 6 begin their

---

[42] Chapters 7 and 8 explain the advantage of collecting these data separately.

[43] Incidentally, the use of summation formulas explains why the spreadsheets do not separate Complements into the categories on which subsystems 4 and 5 are based. I did not decide to break apart these two subsystems until after I had started accumulating data. Adding another column to support subsystem 5 would have meant individually redoing all the automated summation formulas to account for the added column. Given the relative infrequency of these forms, it was easier to do the calculations by hand for them than to change all the automated formulas.

presentation of each group of results with a table derived from the master spreadsheet, which mirrors the totals tab of the individual spreadsheets. These tables are labeled "Overview of Thematization choice in X," where X designates the sample in question.



	Circumstai	Process	Subject	Compleme	Total	
Narrative	185	407	141	13	746	
Projected	142	314	187	110	753	778
Total	327	721	328	123	1499	

Figure 2: Example Spreadsheet of Data

The next step in ascertaining grammatical probabilities is taking this flat count of clause constituents and placing it in paradigmatic perspective. Each type of clause constituent has a "selection expression" associated with it that "represents the paths through the systems visited in the course of" moving through the ideational theme system.[44] Thus, for example, the clause constituent Subject represents the choice -Process in subsystem 1, -Circumstance in subsystem 2, and +Subject in subsystem 3. As a result, based on the data in Figure 2, the frequency of +Subject in Mark's narrative framework is 141; the frequency of -Subject would be the frequencies of clause constituent types that realize choices in subsystem 4 and 5, listed together in Figure 2 as Complements, namely 13. The narrative framework frequency of -Circumstance, in turn, equals the frequencies of +Subject and -Subject combined, 154. An analogous process for the other constituent

[44] Matthiessen et al., *Key Terms*, 188.

types, including breaking down Complements into the choices that are part of subsystems 4 and 5, allows me to reconstruct the frequencies of each choice in the ideational theme system for both the narrative framework and projected discourse. Chapters 4 through 6 report the frequencies with which the author of each text selected the choices in each subsystem, both in absolute terms and as a proportion. They present the results for the narrative framework and projected discourse in separate tables. These tables are labeled "Frequencies of Thematization Choices in X (Y)" where X is the text sample in question and Y is either "Framework" or "Discourse."

Chapters 5 and 6 record the results for multiple texts and, therefore, face the issue of combining these texts into an overall picture of referential and non-referential narrative, respectively. I accomplished this by averaging the relative frequencies observed in each individual sample making up the component of the comparative corpus in question, what is sometimes called the "averages of averages" approach.[45] A problem with this approach is that it "give[s] only a point estimate, i.e. a single number, without any clear indication of how much information we have about the result."[46] That is to say, not all estimates of the grammatical probability are equally precise, so computing their average and computing the average number of times a particular form would occur in 100 instances of that system only provides an indicator of the most common value, not how much spread there is around that central value. As far as I am aware, SFL theory does not take this into consideration, implicitly assuming that the samples will be big enough that the spread will be small. Nevertheless, I will attach notes on the variability of my

[45] Mannila et al., "Quantifying Variation," 342.
[46] Mannila et al., "Quantifying Variation," 342.

results.[47] Chapters 5 and 6 both have two tables reporting these results for referential and non-referential narratives, respectively, labeled "Normed Aggregate Frequencies (Framework) and "Normed Aggregate Frequencies (Discourse)."

Climbing Mt. Genre: Statistical Procedures

The third procedural question regards how one compares maps to each other once the corpus that represents them in place? Consider a craggy mountain. Even if some parts of the mountain are sheer surfaces, others could be relatively flat, allowing a slow, steady descent. Differentiating between the two from a topographical map would involve the shape of the contour lines and the distance between the contour lines. Within the scope of my metaphor, then, comparing Mark to the neighboring genres of referential and non-referential narrative involves the shape of the thematic system in Mark and the shape of the thematic system in the potential analogical genres of referential and non-referential narrative and, particularly, the differences between them. The most relevant analogy will be the one with the smallest difference between its shape and that of Mark.

The quantitative procedures for finding this most relevant analogy have three steps, each of which is the subject of a sub-section below.[48] First, I lay out the procedures for describing the shape of the data I have accumulated, so-called descriptive statistics. In their own right, descriptive statistics explain the procedures for finding the estimates of underlying grammatical probabilities in the comparative corpus and how confident I can be in these estimates. Second, I lay out the procedures for determining what statistical

[47] I discuss the measures of variability below, under the heading "Statistical Procedures for Quantifying Grammatical Systems."

[48] Johnson, "Descriptive," 288.

tools are appropriate for comparing the groups of data I collected to each other. Third, I lay out the procedures for carrying out the comparisons themselves.

Statistical Procedures for Quantifying Grammatical Systems

The mathematical language of statistics already lies behind the notion of corpus linguistics. Likewise, it is part of the background of grammatical probability, which is essentially the mathematical average, or mean, frequency with which a particular feature appears relative to the other option in its system. These and other descriptive statistics serve the purpose of "data reduction" by allowing me to "capture common aspects of a set of observations."[49] Essentially, the process of "drawing the map" described earlier is inductively compiling a frequency distribution.[50] This sub-section describes the procedures for summarizing a distribution. These statistics "are the basis of all quantitative reasoning," and including them in reporting the results of research "is absolutely necessary."[51]

A distribution is definable mathematically in terms of two parameters, namely the typical value and how much the data spreads out around the typical value.[52] Conventionally, the typical value is in the middle of the data set. There are three different ways to measure the typical value. First, the mean, what people colloquially refer to as the average, is defined as the sum of the observations divided by the number of observations.[53] The second measure, called the median, is the middle value of the dataset

[49] Johnson, *Quantitative Methods*, 3; cf. Jenset and McGillivray, *Quantitative Historical Linguistics*, 7.
[50] Cf. Johnson, *Quantitative Methods*, 6–13.
[51] Larson-Hall and Plonsky, "What Gets Reported," 130.
[52] Johnson, "Descriptive," 298.
[53] Gries, "Elementary Statistical Testing," 371.

once they are arranged from least to greatest.[54] The third measure, called the mode, is the most frequent value. Technically, only the third is defined for nominal variables.[55]

The measure of how much the data spreads out around its typical value is called “variance,” defined as the average squared deviation from the mean.[56] Most often, however, the variability of the data is reported in a form that is easier to interpret, known as the “standard deviation” (designated s), which is equal to the square root of the variance.[57] Taken together, the mean (or median) and standard deviation represent the average of the data and the average of differences from the mean. Both these components need to be reported for a complete shape of the data.[58]

According to Daniel Ezra Johnson, while quantitative researchers commonly meet the bare minimum standard of reporting a typical value and a measure of dispersion, typically either the mean and standard deviation or the median and inter-quartile range, they do not go the extra mile and report two other descriptions of the data’s shape, called “skewness” and “kurtosis.”[59] Skewness measures the relationship between the mean and median: sets of data whose mean and median are close together are relatively symmetrical, whereas sets of data for which they are distant exhibit skew. If the mean is higher than the median, the distribution is skewed right; if the mean is lower than the median, the distribution is skewed left.[60] Large amounts of skewness will be visible in a plot of the data.[61] One can also measure it quantitatively as the average cubed difference

---

[54] Gries, “Elementary Statistical Testing,” 371.
[55] Johnson, “Descriptive,” 296.
[56] Woods et al., *Statistics*, 42.
[57] Johnson, “Descriptive,” 299; Woods et al., *Statistics*, 42.
[58] Johnson, “Descriptive,” 298; Larson-Hall and Plonsky, “What Gets Reported,” 128; cf. Eddington, *Statistics*, 10.
[59] Johnson, “Descriptive,” 302.
[60] Johnson, “Descriptive,” 301.
[61] Eddington (*Statistics*, 16) displays this in Figure 2–10.

from the mean divided by the cube of the standard deviation.[62] Kurtosis refers to the "extent to which a distribution has a pointy peak (leptokurtic) or a rounded peak (platykurtotic)."[63] As with skew, large amounts of kurtosis will be obvious from a plot and one can also measure it quantitatively. The formula is very similar to that for measuring skew, replacing the cubing of the skew formula with raising the mean and standard deviation to the fourth power.[64]

The primary reason for reporting these measures of data shape is that their values constrain what statistical tests are valid for a particular set of data.[65] A second use, however, is moving beyond the problems observed above with the process of combining the relative frequencies observed in the various individual texts into normed frequencies for the comparative corpora.[66] If the various texts composing one of the parts of the comparative corpus in fact do represent a coherent group, then they should show similar values for the various measures of similar shape of distribution, i.e. mean, standard deviation, skew, and kurtosis.

These simple descriptive statistics throw some light on how similar the individual texts of the portions of my comparative corpus are to one another, but the light is more diffuse than one would prefer. Essentially, the problem boils down to precisely what qualifies as being similar. Although one could simply set some sort of arbitrary cut-off, identifying cut-off points that are more than simply arbitrary seems difficult because there is no framework for what sort of values are reasonable to expect. Taking a page

---

[62] Johnson, "Descriptive," 302.
[63] Johnson, "Descriptive," 302.
[64] Johnson, "Descriptive," 302.
[65] See the next subsection, "Procedures for Determining What Statistical Tools are Appropriate."
[66] Johnson, "Descriptive," 300.

from parametric inferential statistics, which will be described in detail below, importing such a framework is possible, presuming one is willing to make some assumptions about the nature of the phenomenon in question, e.g. that it conforms to the bell-shaped curve called the "normal distribution."[67] Adopting the normal distribution as a framework automatically provides reference points for skewness and kurtosis because the normal distribution always has a skewness of 0 and a kurtosis of 3. Thus, at a bare minimum, skewness and kurtosis values for two corresponding sets of data need to fall on the same side of these values to qualify as similar. This is, once again, a low bar, for surely one would wish to say something about the magnitude of the difference from the values characteristic of the normal distribution, not just the direction of the difference. Moreover, the normal distribution provides no direct framework for the interpretation of means and standard deviations because the normal distribution can have any mean and standard deviation.

Directly comparing normally-distributed samples with differing means and standard deviations is normally the province of the *t* test, an inferential test described below as part of the discussion of inferential tests.[68] Within the realm of the descriptive statistics on which this section focuses, however, is what is known as a "confidence interval," which consists of an "interval of values around" a point estimate, typically the sample mean, "around which we will assume there is no significant difference" from the point estimate.[69] In other words, the confidence interval is the range of values that would contain a certain percentage of all the sample estimates that could be drawn from a

---

[67] See the next subsection, "Procedures for Determining What Statistical Tools are Appropriate."
[68] See the subsection "Comparing Textual Maps" below.
[69] Gries, *Statistics*, 133.

particular population.[70] Before computing a confidence interval, the researcher has to choose how strict to be about defining a significant difference. The traditional definition of what differs significantly is the most extreme 5 percent, so most confidence intervals (and all of the ones I use here) are 95 percent confidence intervals (100 percent minus the most extreme 5 percent).[71] Beyond the definition of significant difference, the width of a confidence interval depends on a coefficient that further defines the shape of the assumed underlying distribution, and a factor known as the "standard error."[72] Standard error essentially measures how reliably the sample value associated with the confidence interval estimates the typical value in its population.[73] The particulars of calculating standard error and which distribution is used for a coefficient depend on the type of summary statistic with which the confidence interval is associated. Gries gives two examples, namely confidence intervals for the mean and confidence intervals for a proportion.[74] Either of these could potentially work for my project, but I have opted to use confidence intervals for proportions because they allow me to work with absolute, rather than normalized, frequencies, and I have some doubts regarding my normalization procedure.[75]

This being the case, the confidence intervals in the following chapters follow the formula Gries specifies for the confidence interval of a proportion, namely the sample estimate plus or minus the z score (coefficient for the shape of the distribution) times the

[70] Cf. Gries, *Statistics*, 134.

[71] Although stated in converse terms, a 95 percent confidence interval is equivalent to selecting a 5 percent significance level for inferential tests (Gries, *Statistics*, 133, cf. "Statistical Procedures for Comparing Textual Maps" below).

[72] Compare the formula for the confidence interval of a mean (Gries, *Statistics*, 133) and the formula for the confidence interval of a proportion (Gries, *Statistics*, 135).

[73] Gries, *Statistics*, 128.

[74] Gries, *Statistics*, 133–35.

[75] See Chapter 8 under the heading "Lessons Learned."

standard error.[76] In this particular case, the appropriate z-score is the one that cuts off 2.5 percent of the area on either side of the curve (and, thus, a total of 5 percent, leaving the middle 95 percent).[77] Earlier, Gries provides a formula for the standard error of a proportion: (1) multiply the observed proportions for the plus feature and minus feature together, (2) divide by the total number of observations, and (3) take the square root of the result.[78]

Confidence intervals are important because they indicate how well one can generalize from the particular sample in question to the population of which it is a part.[79] This is, of course, the central question for the task at hand, for I am interested in the texts that constitute my comparative corpus only inasmuch as they represent the putative populations from which they spring. Moreover, if the inferential procedures laid out below and carried out in Chapter 7 are to have any meaning, the samples must not only provide reliable estimates of their respective populations, but the populations thus estimated reliably must in fact be those postulated by the procedures. That is, the confidence intervals for the samples labelled as referential narrative need to be consistent with each other, and the samples labelled as non-referential narrative need to be consistent with each other. Ideally, these two aggregated groups would not be consistent with each other, simplifying the task of determining which is more closely aligned with Mark.

---

[76] See formula 22 (Gries, *Statistics*, 135). For ease of reporting, I multiply the results of the formula by 100 to turn proportions back into frequencies.

[77] Gries, *Statistics*, 135.

[78] See formula 19 (Gries, *Statistics*, 129). It is worth noting in this connection that my systems are all binary, so the proportion of the minus feature will be 1 minus the proportion of the plus feature.

[79] Gries, *Statistics*, 133.

How, then, does one determine if two confidence intervals are consistent with each other? If the confidence intervals for two normally-distributed samples of roughly the same size do not overlap, the means of the populations from which the samples spring are significantly different in statistical terms and, thus, the two samples are unlikely to have come from the same population.[80]

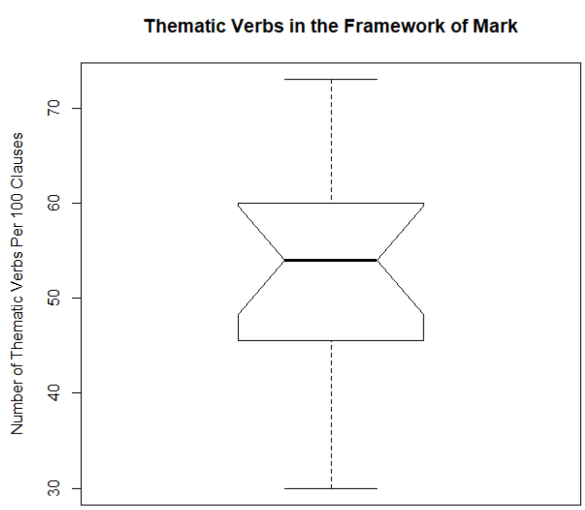
Another way of ascertaining whether individual texts form a coherent group is the box plot.[81] Several components of these plots offer significant information: a thick, horizontal line marks the median of the data sample, the top and bottom lines show the center 50 percent of the data, the end of the whiskers marks the border beyond which any data is mathematically an outlier.[82] Most significantly, however, the notches on the side of the box designate the 95 percent confidence interval, i.e. the edges of the region within which there is a 95 percent chance the true median falls. If these intervals overlap for two samples, the median is probably not significantly different.[83]

An illustration would probably help clarify this discussion. Before I turn to illustrating the problems, however, the figure below, representing the sample taken from Mark's framework on the choice of a verb as ideational theme (subsystem 1), exemplifies what a boxplot should look like. The whiskers show that the extreme data points in this sample are around 30 and 70 selections/100 instances of the system. (Manual inspection shows the precise values are 30 and 73.) The lower 25 percent of the data is between the whisker and the bottom of the box, i.e. between 30 and approximately 45 selections/100

[80] Gries, *Statistics*, 134.
[81] Gries, "Elementary Statistical Testing," 373; Gries, *Statistics*, 126–28; McGill, et al., "Variations," 12.
[82] Gries, "Elementary Statistical Testing," 373; McGill, et al., "Variations," 12.
[83] Gries, "Elementary Statistical Testing," 373; cf. Eddington, *Statistics*, 20.

instances of the system. An additional 25 percent, for a total of 50 percent, lies between 45 selections/100 instances of the system and the median figure of roughly 55 selections/100 instances of the system. An additional 25 percent, for a cumulative total of 75 percent, lies between 55 selections/100 instances of the system and 60 selections/100 instances of the system. The remaining 25 percent is between roughly 60 and 70 selections/100 instances of the system.

Obviously, an even narrower range would be better, but the dispersion of these data is quite reasonable. The data are not quite symmetrical, as the notch testifies: by definition, the notch is symmetrical, being the margin of uncertainty on each side of the median (in this case, roughly 6 selections/100 instances of the system), but the top of the notch seems to reach the top of the box, while there is a small margin between the bottom of the notch and the bottom of the box. Ideally, both sides would have such a margin, indicating both that the data were symmetrical and more confidence in the median of the sample as accurately reflecting the median of the population. Nevertheless, this boxplot reasonably approximates the ideal boxplot.



`

The above illustration of a good boxplot sets the stage for discussing what a less than ideal boxplot indicates about the sample being visualized and, consequently, possible avenues for fixing it. Poor boxplots have triangles protruding from both the upper and lower ends. Two situations can cause these triangles. First, if the confidence interval of the median is wider than the box, the upper border of the box forms the base of a triangle whose third point is the upper boundary of the confidence interval. Second, if either the lower or upper boundary of the box lies sufficiently close to the median that the line for the boundary would be difficult to distinguish from the thick line for the median, the boxplot is drawn with protruding triangles to designate that fact. Either situation entails "low confidence" that the box and, in particular, its estimate of the median as trustworthy estimates of the group the sample is trying to represent.[84] This is almost tautological in the former case, since the degree of confidence in a point estimator, e.g. mean or median, is precisely what a confidence interval measures; in the latter case, the low confidence results from having a full 25 percent of the data, which is the amount that lies between the median and either boundary for the box, being close enough to each other that their values would be difficult to plot.

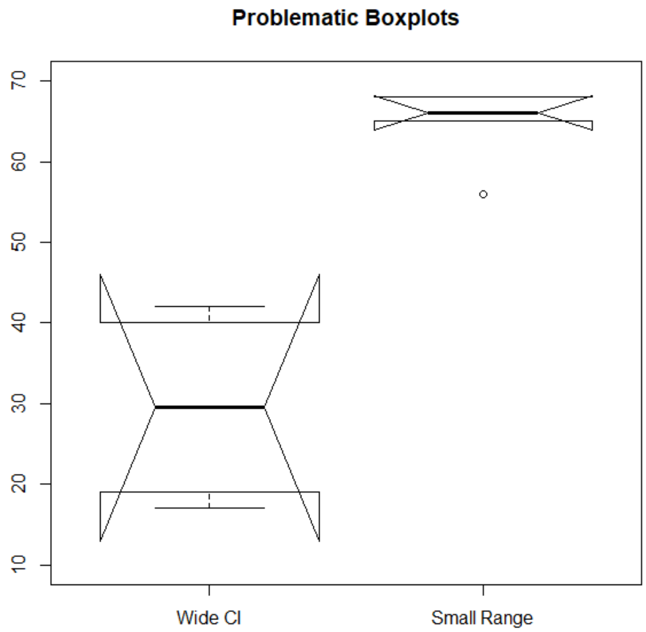
The figure below shows one example of each kind. The boxplot on the left, labelled "Wide CI," visualizes the frequency with which the framework of 1 Maccabees opts for an Adjunct as ideational theme (+Circumstance in subsystem 2) relative to the frequency with which it opts for a category typically realized with a noun (-Circumstance); the boxplot on the right, labelled "Small Range," visualizes the frequency with which the dialogue portions of Judith contain a verb as ideational theme

---

[84] McGill et al., "Variations," 14.

(+Process in subsystem 1) relative to the frequency with which some other type of clause constituent is chosen as ideational theme (-Process). The other output of the R code that generated this figure demonstrates that the problem is indeed the lack of confidence in the sample median as an estimate of the population median: the sample median is 29.5 selections/100 instances of the system, but the believable values for its counterpart in the population ranges from just below 13 selections/100 instances of the system to just above 46 selections/100 instances of the system, both of which lie outside the attested range of values.

Turning to the second case, the other output clearly shows that the problem is not the width of the confidence interval: the median is 66 selections/100 instances of the system, while the upper and lower bound of believable values are just below 64 selections/100 instances of the system and just above 68 selections/100 instances of the system. Rather, the reason that the confidence interval, though narrow, still protrudes beyond the boundaries of the box is that the data exhibit very little variance, meaning that the size of the box is even narrower, as can be seen from the intuitively reasonable value of 56 selections/100 instances of the system qualifying as a mathematical outlier. The height of the box, the distance between the first and third quartiles or the inter-quartile range (IQR) is only 3 selections/100 instances of the system, so it cannot contain a confidence interval roughly 4.25 selections/100 instances of the system wide, despite the latter figure being quite low.



The way forward with either sort of problematic boxplot is the same: extend the sample. In the first case, the formula for the width of a confidence interval calls for multiplying the *t* value for a number of degrees of freedom equal to one less than the sample size by the standard deviation and then dividing by the square root of the sample size.[85] Consulting a *t* table will show that critical values of the *t* distribution get smaller as the sample size increases; likewise, the square root of the sample size increases as the sample size increases. As a result, the width of a confidence interval tends to shrink as the sample size increases: the numerator is getting smaller and the denominator is getting larger. The only way for the confidence interval to get wider with a larger sample would

[85] Gries, *Statistics*, 133. Technically, since this formula uses the *t* distribution, it only works for normally-distributed data. The logic of the caution Gries (*Statistics*, 34) raises with regard to significance testing applies here as well: as with p values, approximating the margin of error around a measure of central tendency "can only be as good as the data's distributional fit to the corresponding function." Nevertheless, the formula clearly and understandably illustrates the role of sample size. Sample size is a major contributor, perhaps an even larger one, to the success of the main non-parametric equivalent for estimating confidence intervals, known as bootstrapping, but that technique is more difficult to explain and implement (cf. Crawley, *Statistics*, 46–50), so I concentrate on parametric confidence intervals here.

be for the standard deviation to increase more steeply than the other factors decreased; as it happens, however, the standard deviation tends to go down as well because the square root of the sample size is the denominator of its calculation as well. Turning to the second kind of problematic boxplot, larger samples provide the opportunity for more values to occur, decreasing the probability of having an IQR smaller than the width of a reasonable confidence interval.

What if one samples the entirety of a document, thus extending the sample as far as possible, and the troublesome triangles have not disappeared? The answer to this question is what differentiates the two types of problematic boxplots. If the confidence interval is still too wide after including the rest of the text, the dispersion of the text must be extreme (in order for the standard deviation to rise faster than the other components of the formula can pull the overall value down), meaning the sample is useless for most comparative purposes because of a low degree of confidence in the accuracy and generalizability of the estimates for measures of central tendency, which are the basis of mathematical comparisons (e.g. means for the *t*-test or ANOVA). On the other hand, if one extends the sample and still finds that the data exhibit so little variability that even a reasonably narrow confidence interval falls outside the box, the situation is quite the opposite: I can place substantial confidence in the measures of central tendency as accurate reflections of those measures in the population and, as a result, most numeric comparisons can proceed without a hitch, although comparison of graphs is unlikely to be helpful.

Procedures for Determining What Statistical Tools Are Appropriate

The shape of this frequency distribution affects what statistical tools are appropriate for comparing grammatical probabilities to Mark's. Certain statistical tools rely on knowing the shape of the data; the most popular of these are the ones that rely on the so-called normal distribution.[86] This being the case, I need to know how well the data conform to the normal distribution. This happens quite often with continuous variables "because a large number of factors cause them to vary, and the sum of a large number of random variables always follows the normal distribution."[87]

The normal distribution is a bell-shaped curve with a single peak in the middle that falls off symmetrically to low points at either end of the graph. This particular distribution has a number of useful mathematical properties. Among these are the ability to quantify exactly how much of the total area under the curve is taken up by each segment of the curve.

Actual data will "never be precisely normal."[88] These sections report both qualitative and quantitative indicators of normality for the data derived from each subsystem, considering the narrative framework and projected discourse separately.[89] They begin with two qualitative indicators, because these graphical displays provide a framework for interpreting some of the quantitative indicators. The first of these is a specific form of bar chart known as a histogram. The horizontal axis of a histogram breaks the data into a series of bars with a width representing the range of the data in that

---

[86] For more information, see "Statistical Procedures for Comparing Textual Maps," below.
[87] Johnson, "Descriptive," 294.
[88] Johnson, "Descriptive," 294.
[89] The next section, titled "Software Assistance for Climbing Mt. Genre," goes over the details of implementing these steps in R, the statistical software package I use.

group and a height representing the number of data points in that group.[90] Histograms of data that have a single peak located near the middle of the graph and fall off more or less symmetrically to each side, i.e. where the bars approximate the shape of a bell curve, support the use of t-tests on that set of data. If the bars' shape differs substantially from this, a t-test will likely return inaccurate results. The second graphical, qualitative indicator, called a quantile-quantile (or Q-Q) plot, is a form of line chart that matches a sample's quartiles against where the quartiles would have fallen for the normal distribution.[91] Histograms hold the size of bins constant and measure the number of observations in a bin; a Q-Q plot holds the number of observations in a bin constant and displays the width of the bins via the spread of the dots.[92]

After presenting and interpreting these two graphical displays, the discussion of normality turns to quantitative measures. With the exception of the standard deviation, which does not help in this regard because the standard deviation only affects the width of the bell curve and, as such, a set of data with any standard deviation could be normally distributed if the standard deviation applies equally on both sides of the mean, the other descriptive statistics described in the preceding sub-section, titled "Statistical Procedures for Quantifying Grammatical Systems," have defined relationships to the normal distribution, allowing them to serve as indicators of the degree to which a particular set of data conforms to the normal distribution. The mean, median, and mode of data perfectly conforming to the normal distribution will be precisely the same, so reporting the measures of central tendency, as these three figures are known, is an important indicator

[90] Johnson, "Descriptive," 292.
[91] According to Johnson ("Descriptive Statistics," 298), quantiles are "the dividing points obtained when you divide the data values into equally sized subsets or bins."
[92] Johnson, "Descriptive Statistics," 298.

of normality. The normal distribution is symmetrical, so its skewness is 0. The kurtosis of the normal distribution is 3, so one should subtract 3 from the computed kurtosis value computed according to the formula described above before comparing the results to the normal distribution.[93] After dividing the skewness and kurtosis numbers by the standard error, a result between -2 and 2 supports conformity to the normal distribution.[94]

Another quantitative measure of normality is to calculate the Pearson correlation coefficient (designated r) between a data sample and the quartile values expected of normal data.[95] This coefficient assigns a value ranging between -1 and 1 to the relationship between two variables, in this case the sample data under consideration and the normal distribution. Strong relationships produce coefficients close to the extremes; a value of 0 would indicate absolutely no relationship. Positive values indicate a direct relationship between the variables (as one increases, the other does as well), whereas a negative value indicates an inverse relationship (as one increases, the other decreases). For my purposes, I consider a correlation coefficient of 0.92 or higher (also written $r \geq 0.92$) to support the use of t-tests.

Neither the qualitative indicators of normality nor the quantitative ones are capable of totally replacing the other. This is true for both practical and theoretical reasons; I focus on the theoretical one here and cover the practical in the next section, since it has to do with coding in R. Qualitative comparison by means of histograms and Q-Q plots is insufficient by itself because the interpretation of these tools is impressionistic and hard to operationalize, thus hindering a quantitative analysis.

---

[93] Johnson, "Descriptive Statistics," 302.
[94] Eddington, *Statistics*, 18.
[95] Johnson, "Descriptive Statistics," 303–05; Eddington, *Statistics*, 27–42.

Quantitative analysis cannot stand on its own, either, because Pearson correlation is unreliable under at least two sorts of situations, both of which I observed over the course of the project.[96] First, if the data have what is known as a bimodal distribution, the calculated value will return a false positive, indicating that the data conform to the normal distribution when they actually represent what is, in some respects, the opposite of a normal distribution.[97] Consulting the histogram flushes out these false positives, however. The histogram of bimodal data will show a pair of peaks located at the extreme values and a valley in the middle. Essentially, these peaks wind up canceling each other out in the process of calculating a Pearson correlation coefficient but the sample has little to no data where the normal distribution predicts the majority of the data would be, so t-tests would be wildly unreliable. Conversely, the presence of a few values wildly out of kilter relative to the others could be enough to cause a false negative.[98] Furthermore, although sets of normal data have measures of central tendency that are identical, having identical measures of central tendency does not guarantee that the data are normal; one also needs to know about the dispersion of the data around this central point. Histograms are particularly useful in this regard, but the vertical spread of the data points on a Q-Q plot is another indicator.

Statistical Procedures for Comparing Textual Maps

Moving on from the descriptive statistics that allow me to quantify the shape of a grammatical system, I proceed to the other main branch of statistics, i.e. inferential statistics, which also impinges on my project as it is the means for determining the

[96] Cf. Gries, “Elementary Statistical Testing,” 377; Eddington, *Statistics*, 32–37.
[97] Cf. Johnson, “Descriptive Statistics,” 305.
[98] Eddington, *Statistics*, 32.

distance between instantiations of a grammatical system. Hypothesis testing, a major focus of this sort of statistics, allows one to make inferences about whether two sets of observations differ because of any reason more significant than random chance.[99] The statistical tools for accomplishing this task are known as significance tests.[100] Over the last century, these tools became "the prevailing statistical procedure for interpreting the findings of quantitative research in most disciplines."[101] The rest of this sub-section is devoted to describing several aspects of significance testing: (1) the logic behind significance tests, (2) a high-level overview of the procedures common to all significance tests, and (3) discussing the peculiarities of the significance tests I used in this project.

The logic undergirding significance testing is often overlooked, even though it is crucial to avoid having one's statistical results devolve into meaninglessness.[102] Strictly speaking, statistical significance tests cannot prove that a hypothesis is true, only that something is unlikely to be true. This being the case, one performs these tests on the so-called "null hypothesis," the converse of the research hypothesis. If the null hypothesis is disproven, then the research hypothesis is likely to be true by process of elimination.

The key factor "is not so much the mastery of the statistical method, which is not overly complex, but is, rather, the selection of the hypotheses."[103] Following S. L. Chow, O'Donnell divides null hypothesis significance tests into two categories on the basis of the nature of the null hypothesis: (1) those where the null hypothesis is "making a decision with regards to whether variation is due to chance" and (2) those where the null

---

[99] Johnson, *Quantitative Methods*, 3.
[100] Cantos, "Use of Linguistic Corpora," 104.
[101] Norris, "Statistical Significance," 101.
[102] Cf., e.g., Eddington, *Statistics*, 19–20; Norris, "Statistical Significance,"; O'Donnell, "Linguistic Fingerprints," 242–45.
[103] O'Donnell, "Linguistic Fingerprints," 244.

hypothesis and the research hypothesis are framed as mutually exclusive "logical premises," such that the law of the excluded middle applies.[104] Confounding these two sorts of null hypotheses is an easy and, unfortunately, common mistake.

What, then, constitutes disproving the null hypothesis? Essentially, one has to decide how unlikely one wants the null hypothesis to be before concluding that the research hypothesis is true. This is a balancing act: setting a low bar for rejecting the null hypothesis runs the risk of adopting the research hypothesis when it is actually false (what statisticians call a "type 1 error"), whereas setting the bar too high can result in failing to reject the null hypothesis when the data actually warrant doing so (a "type 2 error"). Linguists commonly accept a 5% possibility of type 1 error, i.e. they want to be 95% sure that the null hypothesis is false before rejecting it; this is called a significance level of 5%, sometimes written $\alpha=0.05$ or $p < 0.05$.[105]

Null hypotheses typically "assume no variation or lack of change" across a corpus.[106] Thus, properly speaking, significance test results supporting the rejection of a null hypothesis only indicate that a real difference exists between the groups under comparison; they say nothing about the cause(s) or magnitude of that difference.

Since they operate on the same basic logic, significance tests share a general procedure; the differences rest at the nuts-and-bolts level of the precise mathematical operations.[107] This process involves three steps. First, the researcher explicitly sets out a null hypothesis for each test. Second, the researcher plugs the collected data into a formula to compute a value for the test statistic. The formula for this computation is one

---

[104] O'Donnell, "Linguistic Fingerprints," 244.
[105] E.g. Norris, "Statistical Significance," 99.
[106] Cantos, "Use of Linguistic Corpora," 104.
[107] E.g. Norris, "Statistical Significance," 99–100.

of the details that differentiate statistical tests; I will cover this in more detail later. Third, the researcher compares the result of the second step to a set of results known as critical values, found in any statistical textbook. These critical values are simply the values associated with the border of certain common significance levels, so any value greater than the critical value listed for the chosen significance level (5 percent or 0.05, in my case) warrants the researcher rejecting the null hypothesis. For example, one of the significance tests used in this project, which will be discussed in more detail below, has a critical value of 3.841 under certain conditions. If these conditions are met, then a result from step two of 3.842 would cause rejection of the null hypothesis, but a result of 3.840 would not.[108]

Having described the process in general terms, I now turn to describing how it relates to my project specifically. The null hypothesis for each of hypothesis tests follows the same general form. Returning to the above discussion of my project as a variation-based study of corpora, my null hypotheses should fit the following form: "the observations are the result of pure chance, that is, variable x is indistinguishable in corpus A and corpus B."[109] In this structure, x is clause thematization, corpus A is Mark, and corpus B is either referential or non-referential narrative, depending on the particular test in question.

The specifics of steps two and three vary between significance tests, so I need to choose particular significance tests before describing them. Essentially, statistical significance tests fall into two categories, parametric and non-parametric tests. The

[108] If anyone is curious, the test in question is χ^2, and the conditions are a significance level of 5 percent (α=0.05) and 1 degree of freedom.

[109] Cantos, "Use of Linguistic Corpora," 104.

optimal test for a particular set of data depends on how well the data conform to the assumptions of the particular test. Parametric tests get their name because they use parameters, i.e. mathematical properties that are assumed to be true of the data. The information these parameters provide makes parametric tests more powerful: they will detect a statistically significant pattern from a smaller sample than a non-parametric test would, if the parameters are actually true.[110] If the assumptions are inaccurate, however, the test is worthless. Conversely, it takes a larger sample for a non-parametric test to find a given pattern, but a sufficiently large sample guarantees an accurate result.

The best path forward seemed to be combining the two. Chapter 7 uses a non-parametric test called χ^2 analysis and, if the descriptive statistics computed above justify it, a parametric test called the t-test as well. Though χ^2 analysis makes fewer assumptions than a parametric test, there are still two important ones: (1) the data need to be independent of each other and (2) the researcher needs sufficient data for every cell to have a value of at least 5.[111] The ease with which one can demonstrate that a set of data meets these assumptions has made χ^2 analysis particularly popular in linguistic circles. [112] The largest advantage of χ^2 analysis is that, unlike most statistical significance tests, it works with absolute frequencies, instead of proportions, meaning that it can handle what are called nominal variables. Nominal variables are variables like my data, the frequency

---

[110] Eddington, *Statistics*, 37.

[111] Cf. Butler, *Statistics*, 112–23; Eddington, *Statistics*, 50. Libby notes that more complex multivariate methods are becoming more common in statistical linguistics ("Disentangling," 38). However, the context of Libby's comment seems to indicate that he is discussing those who are expecting the statistical analysis itself to bear the brunt of the analysis. I, on the other hand, am only using the statistical analysis to observe whether or not any observed deviations in grammatical probabilities are sufficiently large to require the invocation of a more complex hypothesis than random chance in order to explain them. X^2 analysis should be sufficient for this more limited task. In fact, testing the "goodness of fit" between a theoretical model and a set of observations is one of the classical applications for such analysis (Butler, *Statistics*, 114–18).

[112] Butler, *Statistics*, 74–75, 112; cf. Libby, "Disentangling," 13n.43.

with which a particular kind of clause constituent appears first in a primary clause. A particular clause constituent either is or is not functioning as the subject of a particular clause; something cannot be twice as Subject as another thing, nor is there some implicit zero where there is no Subject-ness. Thus, rather than being a "ratio" or "ordinal" variable, frequency data are a "nominal" variable, and mathematical operations that assume another type of variable, such as the sample means that are part of a t-test, are technically inadmissible, though research has shown that the test is relatively robust to the violation of this assumption.[113]

By contrast, the "outcomes of" parametric tests, like the t-test, "can be dramatically affected by data that do not conform to assumptions such as normality of distributions," among others.[114] According to one applied linguist familiar with quantitative approaches, failure to consider whether or not data fit the assumptions of the test(s) employed "suggests that findings related to statistical significance are suspect at best" for the majority of second-language research.[115] I should perhaps point out, in fairness to both Norris and those to whom he refers, that his statement is grounded, as he recognizes, purely on whether the researchers in question explicitly mentioned that they considered the assumptions; obviously, Norris has no way of evaluating how many of these scholars considered the assumptions but did not choose to include that fact in their research report. Recognizing this, however, provides all the more reason to include a report of such consideration, lest one's reader assume it was not considered and consequently dismiss the value of one's statistical conclusions.[116] Chapters 4, 5, and 6,

[113] Cf., e.g., Johnson, *Quantitative Methods*, 4–5.
[114] Norris, "Statistical Significance," 105; cf. Eddington, *Statistics*, 55–56.
[115] Norris, "Statistical Significance," 105.
[116] Eddington, *Statistics*, 56; Larson-Hall and Plonsky, "What Gets Reported," 130.

the chapters devoted to Mark, referential narrative, and non-referential narratives, respectively, each have a section evaluating the normality of the data for precisely this reason.

Carrying out χ^2 analysis requires several pieces of information.[117] First of all, in terms of a qualitative framework, one needs to know what categorical variables are to be compared, and the values these variables can have. In my case, the categories are normally Mark and either referential or non-referential narrative. The clauses of which each of these categories consist all select from the ideational theme system described above, and within each subsystem that choice is either +X or -X, where X is the clause constituent type associated with the subsystem in question. This means the “contingency table,” which forms the basis of χ^2 analysis, has two rows (one for Mark and the other for the component of the comparative corpus with which the particular test is concerned) and two columns (one for +X and one for -X). A 2x2 table has four cells, and within each of them I need both an observed and expected frequency. The observed part is easy: I just plug in the number for the category that results from the process described above under “Drawing the Maps.” The expected frequency is slightly more complicated because it depends on the null hypothesis. Above, I stated that my null hypothesis assumes no difference between the two categories under consideration. If this null hypothesis is true, the percentage of times each choice is selected should be the same for both documents. I can compute this value based on the total values for both documents. Multiplying this proportion by the observed value in each cell generates the expected frequency. Once the expected frequencies are in place, I plug the observed and expected frequencies into the

[117] Cf. Cantos, “Use of Linguistic Corpora,” 116.

χ^2 formula and generate the χ^2 value for that cell and, once all four are done, aggregate them to produce a total χ^2 value for the table. This is the value I compare with the critical values of the χ^2 statistic. This study only has 2x2 tables, so the degrees of freedom will always be 1. This being the case, and having selected $\alpha=0.05$ as my significance level earlier in the chapter, the critical value for all my χ^2 tests is 3.841.

I also perform t-tests on my data when I can justify them.[118] In contrast to χ^2 analysis, the t-test works on the basis of sample proportions, i.e. relative frequencies. This means that the frequencies observed in texts of different lengths need to be put on a common basis for comparison, a process referred to as normalization.[119] Consider two hypothetical scenarios illustrating the necessity of normalization. First, consider a situation where one observed 10 instances of a particular linguistic phenomenon in a text of 100 words and 10 instances of the same phenomenon in a text of 1000 words; in such a case, the absolute frequencies are equal but the proportion is clearly different. Conversely, consider a situation where one observed 20 instances in a text of 100 words and 200 instances in a text of 1000 words; in such a case, the absolute frequencies are different but the relative frequency is absolutely the same.[120] For this study, rather than using a particular number of words as the common denominator between two texts of uneven length, the proper basis is a particular number of instances of the subsystem under consideration. I chose 100 instances of a system as the basis for my figures.[121]

[118] Cf. Eddington, *Statistics*, 53–68.
[119] Cf., e.g., Cantos, “Use of Linguistic Corpora,” 108–9; Tardy, “Genre,” 67n.1.
[120] Cf. Woods et al., *Statistics*, 9.
[121] Chapter 8, in the section labeled “Lessons Learned,” offers some reflections on why I would choose a smaller basis moving forward.

Once the normalized frequencies are in place, the test consists of comparing the chances that the means of the two distributions (Mark and which ever component of the comparative corpus is under consideration) are the same.[122] As mentioned above, the t-test assumes that the two samples conform reasonably well to the normal distribution. Another assumption of the default form of the t-test, the so-called "student's t-test," is that the spread around the mean in both data sets is approximately equal. The advantage of such a test is that it is easier to compute by hand. Given that I am using statistical software to perform the calculations anyway, however, I have no reason not to use the so-called "Welch's t-test," which makes no such assumption.[123]

Regardless of the form of t-test used, ascertaining the value of the test statistic depends on identifying whether the test in question has one or two tails. This distinction refers to how the chance of type 1 error is distributed: a one-tailed test places the rejection region all on one side of the bell curve, whereas a two-tailed test divides it equally between left and right sides. The hypotheses for a one-tailed test are directional (e.g. "the mean of sample X's population will be greater than the mean of sample Y's population); the hypotheses for a two-tailed test are non-directional (e.g. "the mean of sample X's population differs from the mean of sample Y's population"). Given that I only care whether the data from Mark and the data from one of the components of the comparative corpus indicate that they come from groups of texts with different characteristics, all my tests are two-tailed. The null hypotheses are all of the following

[122] Eddington (*Statistics*, 53) illustrates this graphically.

[123] One wrinkle Welch's t-test does throw into the procedure is that I cannot state beforehand what the critical value of the test statistic will be. The degrees of freedom for Welch's t-test are calculated separately for each test based on the characteristics of the samples involved, rather than being a constant value for all tests, like, e.g., my χ^2 analyses, which I know will always have 1 degree of freedom. This being the case, I cannot state a general critical value for all my t-tests the way I did above. Chapter 7's t-tests each have a footnote referencing a table of critical values, instead.

form: the means of the populations from which Mark and X are derived are equal, where X is the component of the comparative corpus in question.

One further step remains after running the significance tests themselves. As discussed earlier, null hypothesis significance testing, at least as I am practicing it here, is not equipped to reach conclusions about the cause or magnitude of any significant differences it finds. While a significant result from a hypothesis test indicates a significant difference, a non-significant result does not necessarily indicate there is no significant difference. First, NHST procedure constrains the probability of false positives ("type 1 error"), which of necessity increases the probability of false negatives ("type 2 error"). Type 2 error is particularly likely in cases where the p-value associated with a hypothesis test is close to the cut-off value; as the p-value increases, type 2 error becomes less likely. Second, sample size profoundly affects how easily a researcher can reject the null hypothesis. Consider the formula for a χ^2 test: the difference between observed and expected values is squared before dividing by the expected value, meaning that as total sample size increases the numerator will grow faster than the denominator and, thus, reaching the cut-off value of 3.84 will be easier. The same is true for the t-test, though that case is more difficult to illustrate. The number of primary clauses in Mark, referential narratives, and non-referential narratives are not all equal, so I need some sort of common ground that takes sample size out of the equation before comparing significance test results will be meaningful.

The proper tools for determining how significant a significant finding is are measures of effect size because sample size does not affect them. As Larson-Hall and Plonsky recommend, I will report these measures for all results, regardless of whether the test result reached statistical significance, and the majority of the interpretive weight rests on them.[124] Both tests of statistical significance used in this project have at least one such a measure. The measures of effect size for χ^2 analysis are standardized residuals and Cramer's V; the measure for the effect size of a t-test is Cohen's d.

The residuals of a χ^2 table, i.e. the values each cell contributes to the total value of the test statistic, can provide an initial indicator of the cause by showing which cell is out of step with the others. Even small differences between the observed and expected frequencies eventually add up, however, so this use of the residuals requires dividing them by the square root of the number of observations to take the effects of sample size into account.[125] This process is known as standardizing the residual.[126] If the resulting number is greater than 1.96, then one can conclude with 95 percent confidence that the cause of significant difference involves the cell in question.[127] The usefulness of this particular measure is limited, however, so I have chosen to report the results of Cramer's V instead.

Cramer's V is what is known as a test of association. To compute it, one divides the value of the χ^2 test statistic by the total number of observations and then takes the square root of the result.[128] The results of this test range between 0 and 1, and Gries gives the following "frequently-used classification" brackets for interpreting Cramer's V: a result less than 0.1 indicates a negligible effect, a result between 0.1 and 0.3 indicates a "small effect," a result between 0.3 and 0.5 indicates a "medium effect," and any result

[124] Larson-Hall and Plonsky, "What Gets Reported," 128–29.
[125] Eddington, *Statistics*, 48.
[126] Eddington, *Statistics*, 48.
[127] Eddington, *Statistics*, 48.
[128] Gries, "Elementary Statistical Testing," 370.

over 0.5 indicates a "large effect."[129] The discussion of effect sizes in Chapter 7 uses these brackets for interpreting the relevant tests.

The measure of effect size for the results of a t-test is called Cohen's d.[130] Essentially, this test measures how many standard deviations lie between the means the t-test compared. The standard conventions for interpreting this statistic are: a difference in means of less than 0.2 standard deviations is a "negligible effect," a difference in means of between 0.2 and 0.5 standard deviations is a "small effect," a difference in means of between 0.5 and 0.8 standard deviations is a "medium effect," and a difference in means of more than 0.8 standard deviations is a "large effect."[131] The discussion of effect size in Chapter 7 uses these brackets for interpreting the relevant tests.

At this point, perhaps an illustration of a linguistic application of effect size statistics would be helpful. Douglas Biber and Jesse Egbert's register analysis of the Internet uses Cohen's *d* as a measure of "keyness" for features of lexicogrammar.[132] They characterize the measure as the difference between "the mean rate of occurrence for a linguistic feature in the target corpus" and "the mean rate of occurrence for the same feature in the reference corpus," divided by the "pooled standard deviation" so that "the result [is] on a standardized scale."[133] The purpose of using this measure is "to assign higher rankings to lexico-grammatical features that are used much more in the target (large positive *d* values) and much less in the target (large negative *d* values)."[134]

---

[129] Gries, "Elementary Statistical Testing," 371n5.

[130] Eddington, *Statistics*, 54.

[131] As a simplification, these brackets are the absolute values of the brackets Eddington (*Statistics*, 54) suggests. I am using the absolute values because the sign associated with Cohen's d only indicates the direction of the difference, i.e. whether the mean of the first group is higher or lower than that of the second. My hypotheses are non-directional and, as such, the sign is irrelevant.

[132] Biber and Egbert, *Register*, 23–25.

[133] Biber and Egbert, *Register*, 24.

[134] Biber and Egbert, *Register*, 24.

Applying this to my project, my reference corpora are the categories of referential and non-referential narrative, and the target corpus is Mark. Thus, Cohen's d, and by extension Cramer's V as the analogous test for χ^2 tests, show how different the respective sub-corpora are from Mark.

Software Assistance for Climbing Mt. Genre

This section covers the details of how I used a statistical software package to facilitate the procedures described above. The statistical software package in question is called R. While other statistical software tends to be proprietary and quite expensive, R is free and its source code is freely available, allowing it to "become the de-facto tool in the field of statistics and is often cited as being amongst the Top-20 used programming languages in the world."[135] Had I learned of R's existence earlier, I would have organized the entire project differently, but even using it only in the latter stages of the project has still been invaluable.[136]

The rationale for including a discussion of R here is two-fold: (1) some readers may be interested in greater detail concerning how the numbers reported in the ensuing chapters are calculated than the preceding overview of procedure provided and (2) discussions of the role of R is a burgeoning topic in corpus linguistics and related fields but, to my knowledge, no one has discussed its potential for quantitative study of Hellenistic Greek in general and the New Testament in particular.[137] This section is a step towards filling in that lacuna. It consists of four subsections. The first describes the

---

[135] Arnold and Tilton, *Humanities Data*, 3.

[136] More details on the effect knowing of R could have had on the organization of the project can be found in chapter 8, under the heading "Lessons Learned."

[137] Cf., e.g., Arnold and Tilton, *Humanities Data*, passim; Gries, "Elementary Statistical Testing," 361–81; Jockers, *Text Analysis*, passim.

advantages of R, its background, how to get it, and basic concepts of R syntax so that the reader can hopefully follow the snippets of code and programming jargon that are an inescapable part of the latter three subsections, each of which corresponds to one of the subsections in the preceding section.

Software Assistance for Climbing Mt. Genre: Prolegomena

R offers a number of advantages for a project like this. Perhaps the foremost among these is the community of programmers who have taken advantage of R's open-source framework and written code to implement a vast array of tools for analyzing data.[138] These tools save researchers time and effort reinventing the wheel, and they also vastly reduce the amount of programming knowledge needed to use R.[139] Rather than needing a thorough grasp of both the calculations involved in the tool being used and the programming knowledge to tell the computer how to carry them out, researchers simply need to understand the format in which the code expects to receive the data, which is typically covered in accompanying documentation, and have enough programming knowledge to understand the process of loading packages and telling R the desired piece of code, called a function, that one wants to run.[140] Another selling point of R is its ability to produce pictures that paint the proverbial one thousand words: R is capable of producing professional-looking charts and graphs that communicate the latent

---

[138] Cf. Arnold et al., "Beyond Lexical Frequencies," 713. Desagulier (*Corpus,* 14) lists several online forums and code repositories in which one can find help, or even pre-written code, to implement a wide variety of tasks.

[139] Arnold et al., "Beyond Lexical Frequencies," 712–13.

[140] Arnold et al. ("Beyond Lexical Frequencies," 715) divide perspectives on the use of R in digital humanities into those who treat "R as a general purpose programming language" and those who treat it "as a collection of useful packages." I am following the latter perspective here. For the purposes of this section, I am simply reporting which packages I used and the R code to call them so that the reader can follow up on my calculations, if they wish.

relationships within the data both to researcher him- or herself and the readers with whom he or she wishes to communicate the results.

R is freely available for a variety of hardware platforms at http://cran.r-project.org.[141] This is the same site that serves as a repository for the packages referred to above. Once the software is downloaded, one can install packages from inside the software itself. The majority of the code I needed was available within the base packages that come with R; the code to install the two extra packages I used is: “install.packages (“moments”)” and “install.packages (“effsize”). After installing them on one’s hard drive, the next step is to tell R that they are to be loaded into this particular project by typing these two lines of code: “library(moments)” and “library(effsize).” Once these four lines have run, the functions in these two packages are available; they will be described below at the relevant points below.

Researchers with a background in computer programming, even one as rudimentary as my own, will probably find the basics of using R fairly intuitive, but I will summarize them here.[142] The description of R as a programming language is descriptive, but its vocabulary and syntax are far more restrictive than natural languages. At the most basic level, two entities form the bedrock of R, namely functions and data structures. Functions are groups of code that do something and then return a result; data structures are defined ways of giving functions the input and receiving the output from them. The input(s) for a function, called argument(s), are found in data structures; R expects to see

[141] Arnold and Tilton, *Humanities Data*, 5; cf. Desgaulier, *Corpus*, 16. The current version is 3.6.2; I have version 3.5.3 installed.

[142] This paragraph builds on Arnold and Tilton, *Humanities Data* (7–24).

these inside parentheses immediately following the function's name.[143] Output from a function can also be stored in a data structure; the operator to tell R to do this is a less than sign followed by a minus sign, basically an arrow pointing from the function to the name of the data structure in which you want to store the output.[144]

Sometimes one is only interested in part of the output for a function; one does this by placing a $ between the argument's name and the name of the component one wishes to keep. The main reason to do this is if one needs to pass this data on to another function that expects to receive its data in a specific form. One recurrent example of this in my project is needing to compute a Pearson correlation coefficient between each group of data and the normal distribution. The particular form of the function to compute a Pearson correlation coefficient that I am using requires that the input data be as vectors, essentially a series of numbers, but the function from which I get the data for the normal distribution part of the correlation outputs a number of vectors, so I have to specify which one I want.

Thus, a full function call potentially involves, moving from left to right: (1) the name of a data structure in which to save the output of the function, if this is present, one also needs the arrow operator, (2) the function name, and (3) the arguments the function requires, which may involve the $ operator to specify what part to save). In the following example, process_frame_Mark_norm is the name of a data structure that needs to be in vector form so that it can then serve as part of the calculation of a Pearson correlation coefficient, qqnorm is the function that tells R to draw a quantile-quantile plot, discussed

[143] R's help system specifies the arguments required for each function and the type of data structure it expects each argument to be. To access this, simply type the function name, preceded by a question mark.

[144] One uses the same operator to assign values to a data structure in the first place.

above under “Procedures for Determining What Statistical Tests are Appropriate,” and process_frame_Mark is a vector data structure containing the normalized frequencies by chapter for choosing a verb as ideational theme, i.e. +Process in subsystem 1. The line of code is: “process_frame_Mark_norm <- qqnorm(process_frame_Mark)$x.” To illustrate how the result would then be passed to another function, the ensuing code to compute a Pearson correlation coefficient would be: “cor (process_frame_Mark, process_frame_Mark_norm).”

The ability to produce compelling data visualizations has been a hallmark of R and its predecessor since the very beginning. The predecessor of R, oddly enough called S, was a proprietary package developed at Bell Labs as John Tukey, a scientist who worked there and at Princeton University, was developing a research approach known as “explanatory data analysis” that “prioritizes studying data directly in order to generate hypotheses and ascertain general trends prior to, and often in lieu of, formal statistical modeling” by means of summary statistics and graphs.[145] As one would expect, Tukey and his coworkers ensured that the software package they produced was well-suited to the research methodology they favored. As a result, S is particularly well-suited for pursuing explanatory data analysis; R, the completely open-source version of S, is no different.[146] This means that R’s resources for descriptive statistics and graphs are particularly well-developed, and both these aspects figure into the first two subsections below, which explain the code with which I implemented the descriptions of thematization within the texts I examined. Moreover, despite devoting considerable attention to the tasks involved in exploratory approaches, R does not neglect more

[145] Arnold and Tilton, *Humanities Data*, 3.
[146] Arnold and Tilton, *Humanities Data*, 4.

traditional hypothesis-testing approaches, and the third subsection explains the code necessary to implement that portion of the project.

Software Assistance for Quantifying Grammatical Systems

I discussed several types of descriptive statistics in the previous section under the heading "Quantifying Grammatical Systems." These include the three measures of central tendency (mean, median, and mode), the standard deviation, the skewness, and the kurtosis. This subsection describes the code necessary to implement these statistics in R.

While there is no built-in function to calculate the mode of a set of data, there are functions for the other two measures of central tendency. The functions for both mean and median require one argument, namely the name of the data structure that has the data whose mean and median R is to find. Logically enough, the functions are named mean and median, so the following lines of code compute these statistics for a dataset called x: "mean(x); median (x)." Finding the mode manually within R is possible, though somewhat tedious, depending on the length of the data. If the user types the name of a data structure without any other commands on that line, R will list the values in that data structure. Then finding the mode involves counting the number of times a particular value appears.

Similarly, the functions for standard deviation**,** skewness, and kurtosis are all named intuitively and take one argument, namely the label of the data structure containing the data whose shape is in question. Unlike the functions for the measures of central tendency and standard deviation, however, the functions for skewness and kurtosis are not included in base R. They are located in a package called "moments," which must be installed and loaded using the code described above. Assuming the moments package is in memory, the following code will calculate the standard deviation, skewness, and kurtosis of a group of data called x: “sd(x); skewness(x); kurtosis(x).”

I used several R functions as I compiled the results reported in chapters 4 through 6. The line of code to compute a confidence interval for a proportion is: “prop.test (success, total, correct=F”, where success is the number of instances observed of the realization of the plus choice in a given subsystem, total is the number of choices made in the subsystem, i.e. both plus and minus choices, and “F” is R’s shorthand for “FALSE.”[147] The value returned from this function is saved in a variable, and then one can report the confidence interval by running another line of code that asks for the “conf.int” component of the variable to which you saved the results of prop.test.[148] Drawing an example from the results reported in the next chapter, I observed 409 instances of thematic verbs, the realization of +Process in subsystem 1, within the narrative framework of Mark, out of a total of 746 choices. The code to compute the confidence interval for the observed percentage would be: Mkframesub1ci<-prop.test (409, 746, correct=F) and then, after hitting Enter to run that line of code, another line: Mkframesub1ci$conf.int. Hitting Enter again displays two numbers, which are the lower and upper bounds of the interval. In this case, I can be 95 percent sure that the true value of the chance of having the verb be the ideational theme of primary clauses in the narrative framework of texts of Mark’s register is between 51 and 58 percent, which matches the observed frequency of 55 percent quite well. As it turns out, I wound up using this function as a cross-check for manual calculation of the range of the confidence interval because it only returns the boundaries of the confidence interval, whereas I

[147] Gries, “Basic Significance Testing,” 327.
[148] Gries, “Basic Significance Testing,” 327.

wanted to use the succinct notation that reports the mean and the distance from the mean to the boundaries. The following code computes this distance manually, given a proportion p and sample size n: “1.96*sqrt(p*(1-p)/n).”

The boxplots described in the corresponding subsection of the previous section are also relatively easy to produce within R as a result of its focus on visualizing data. As a matter of fact, after running the code explained in the rest of the paragraph for each set of data, I copied the boxplot that appeared in R and pasted it directly into the word-processing file of the relevant chapter. Once again, the function is intuitively named. Like the functions for basic descriptive statistics explained above, the boxplot function requires the names of the data structures containing the data to be plotted as arguments; in contrast to those functions, however, it also requires several other arguments. Only one of those not concerned with labels needed to be reset from its default value for my plots, namely “notch.” R’s default is to draw boxplots without the notches indicating confidence intervals, but said intervals are a large part of the reason I am using boxplots in the first place, so I need to set “notch=T,” which is R’s shorthand for “TRUE.” As for labelling parameters, “main” sets the title of the whole plot, “xlab” and “ylab” label the x and y axes, respectively, and “names” takes a vector of character strings whose length matches the number of data structures that were passed as data. The individual elements of this vector then serve as labels for the individual boxes in the plot. Lastly, saving the output of the boxplot function to a variable allows one to retrieve components of this output later, which comes in particularly handy when the exact boundaries of a confidence interval are unclear on the plot due to scale. Putting all this together, a full function call to create a side-by-side boxplot of two groups of data named a and b would look like this: example.bplot<-boxplot(a, b, notch=T, main="Title", xlab= "X-axis Label", ylab="Y-axis Label", names=c("Label for A Box", "Label for B Box")).[149]

This subsection has laid out the code necessary to calculate various sorts of descriptive statistics. Chapters 4 through 6 use these statistics in several ways, including evaluating how similar the individual samples I have grouped together are to one another and how similar they are to the normal distribution. The raw data behind these discussions comes from executing the code described above for each sample. In particular, R generated the boxplots that appear in Chapters 5 and 6.

Software Assistance for Determining What Statistical Tests are Appropriate

This subsection explains the code necessary to implement the procedures for testing the normality of a set of data, set out above under "Procedures for Determining What Statistical Tests are Appropriate. Since the preceding subsection already explained the code for calculating descriptive statistics, I will focus here on the code for other parts of the procedure, namely histograms, Q-Q plots, and Pearson's correlation coefficients.

The hist function in R draws histograms. The only required argument is the name of the data structure containing the data the histogram is visualizing. A number of optional arguments are available to customize the look of the histogram, but I stuck with the default values. The function call for a set of data named x is: "hist(x)."

The qqnorm function in R draws Quantile-Quantile plots that compare a data sample to the normal distribution. It takes the sample to compare to the normal

[149] One pitfall I ran into while inputting this code, since I was working back and forth between R and the electronic files of the dissertation itself, was automatically placing the commas inside the quotation marks, in accordance with the dissertation's style. I had to overcome this tendency. If the commas are inside the quotation marks, R treats them as part of the character string being assigned to the parameter and does not process it properly.

distribution as a required argument. It also offers a variety of other arguments related to labels. One particular component of the output for this function, namely the x component of the points, needs to be saved to a variable because it is required for calculating the Pearson's correlation coefficients described below. In addition to the plot itself, the qqline function draws a reference line that marks where the points would be if the sample conformed to the normal distribution perfectly. Once again, the only required argument for the qqline function is the data structure containing the data to be compared with the normal distribution. The following line of code will generate a Q-Q plot for a sample named a, save the x coordinates of its points to a variable called "ex.norm" for further use, and draw a reference line: "ex.norm<-qqnorm(a)$x; qqline(a)."

The cor function calculates, among other things, Pearson's correlation coefficient. Pearson is the default method, so the only arguments passed to this function are the names of the data structures containing the two samples being compared. In the case of testing a sample for conformity to the normal distribution, one of these will be the sample passed as an argument to the corresponding call to the qqnorm function, while the other will be the output saved from that function call. For instance, the following line of code will calculate Pearson's correlation coefficient for the Q-Q plot from the sample code in the preceding paragraph: "cor(a, ex.norm)."

This subsection has laid out the code to execute the normality testing procedure laid out in the preceding section. Chapters 4 through 6 each contain a section reporting how well the various samples with which the chapter in question is concerned conform to the normal distribution. The raw data behind these discussions comes from executing the code set out here. In fact, the histograms and Q-Q plots in those chapters are copied directly from R's plot window.

Software Assistance for Comparing Textual Maps

This subsection explains the code necessary to implement the procedures for implanting the statistical tests described above under "Statistical Procedures for Comparing Textual Maps." This involves code to execute both χ^2 analysis and t-tests, along with their respective measures of effect size (Cramer's V and Cohen's d).

R handles χ^2 analysis with a flexible function, chisq.test, that can handle both multiple forms of χ^2 analysis and can take its input in several forms. As a result, it takes a variety of arguments with which a user can specify what sort of χ^2 test he or she wishes to run and how R should handle the input. Chapter 7's χ^2 tests are all tests of independence and use calculated p-values, so the relevant arguments are how the input is passed to the function and whether continuity correction is used. I passed the data to the function as a 2x2 matrix, so the function only needs its name, rather than two names as would be the case if I passed the rows individually. I also chose to use continuity correction, even though its effect on some of my analyses would be quite small, because it is recommended for 2x2 tables with small frequencies, such as the more delicate subsystems in my analyses. Lastly, saving the output of the function allows one to execute other steps more easily. Thus, a full function call for a χ^2 test, assuming a contingency table called contig.table is already in memory, looks like this: "ex.chi2<-chisq.test (contig.table, correct=T)." The other arguments do not apply to tests like mine.

Returning to the form in which the data are passed to chisq.test, I chose to pass the data to the function as a matrix, for that form most closely resembles the contingency tables with which I was familiar from having run χ^2 analysis manually in earlier work. One complication this decision entails is producing the contingency table in the first place. To do so, I used the rbind function, which takes two or more vectors as arguments and binds them together as a table. The initial r in the name signifies that this binding takes place row-wise. That is, the first vector constitutes the first row, the second vector constitutes the second row, etc. I needed a 2x2 matrix, so I passed two vectors to rbind. I could have created these vectors first by assigning them to variables and then passing these variables as arguments to rbind. In the interests of efficient coding, however, I instead used the c function to create the vectors within the call to rbind itself, which both saves two lines of code and reduces the number of objects stored in memory by two per table or forty over the course of the tests reported in Chapter 7. The numbers being combined came from the total observed frequencies recorded for the various choices in the relevant tables of Chapters 4 through 6. For instance, for a test comparing Mark's instantiation of subsystem 1 in his framework to the instantiation of subsystem 1 in the framework of referential narratives, the first vector would consist of the total frequency of the choice +Process in Mark's framework and the total frequency of the choice -Process in Mark's framework. Proceeding with this logic, a full function call to generate a contingency table with observed frequencies of a, b, c, and d would look like this: "contig.table<-rbind(c(a,b), c(c,d))."

The last issue to cover with regard to χ^2 analysis is the output. Running the chisq.test function provides only minimal visible output (normally just the value of the test statistic, the degrees of freedom, and the p-value), but more is available in the background by saving the output to a variable and using the $ operator. The frequencies expected under the null hypothesis are available by this means ($expected), as are the standardized residuals that help explain precisely where the significant deviation lies when a test indicates such ($stdres). Additionally, the material printed on-screen by the standard output is also available this way so that it can be included in further calculations. This is particularly useful for the value of the test statistic ($statistic) and the observed counts ($observed), both of which figure into the discussion of Cramer's V below.

As previously discussed, the measure of effect size I am reporting for my χ^2 tests is Cramer's V. As far as I have been able to determine, R has no built-in function to compute this statistic. However, at least for the 2x2 tables I am using, it is simple enough to calculate directly on the command line, using R as a calculator. Cramer's V for a 2x2 table is defined as the square root of the test statistic divided by the total number of observations. In R terms, the following line of code will output the value of Cramer's V for a χ^2 test saved as data.X2: "sqrt((data.X2$statistic/sum(data.X2$observed)))."

As with χ^2 analysis, R uses one function, in this case t.test, to cover a variety of different tests (one sample, two samples, paired, etc.) by allowing the user to specify the particular type of t-test in which he or she is interested via the particular arguments passed. The t-tests in this project happen to match the default settings for the function, so I only need to provide two arguments explicitly, specifically the name of the data structure containing the two samples to be compared. As such, the following line of code is a complete function call to run a t-test comparing two samples named a and b: "t.test (a,b)."

An R function to compute Cohen's d is available in the "effsize" package. This function, cohen.d, takes a variety of arguments. Only two of them, namely the names of the data structure containing the two samples whose difference is being compared, apply here; I used the default values for the others. Assuming the package is already in memory, the following line of code calculates Cohen's d for samples named a and b: "cohen.d (a,b)."

Summary and Structure of What Follows

This concludes the preliminary portion of this project. Having described how my project fits into the history of the discipline (Chapter 1), the theoretical framework that motivated the investigation (Chapter 2), and the procedure by which I carried out the investigation (Chapter 3), I now turn to describing the procedure by which I will present the results of that investigation.

Chapters 4 through 6, which report the ideational theme choices observed in Mark and the comparative corpus, share several structural similarities. Each begins with a section detailing problematic choices I encountered in annotating the texts. They all have two sections presenting the results of the procedure described above under the heading "Drawing the Maps," one devoted to the narrative framework and one devoted to projected discourse. They all have a section reporting the descriptive statistics laid out above under "Statistical Procedures for Quantifying Grammatical Systems," along with a qualitative interpretation of these data. The last major section of each chapter details what statistical tools are appropriate for each group of data, following the procedure laid out above under "Procedures for Determining What Statistical Tools Are Appropriate."

Chapter 4 has one component that differs from the other two and, conversely, they have an element that was not needed for the data from Mark. Specifically, the data from Mark were the best choice for ascertaining whether I needed to control for a clause's status as part of the narrative framework or projected discourse because it was the longest text where I used exhaustive enumeration, rather than simply testing a sample. Conversely, both sections of the comparative corpus consist of multiple documents, so the chapters devoted to them spend some time discussing the procedure for combining the data from these documents to produce an overall estimate of the grammatical probability of the various clause constituent types being chosen as ideational theme. Chapter 4 did not need this component because I was not aggregating data from multiple documents.

Chapter 7 compares Mark's distribution of theme choices to those found in the previous two chapters on the components of the corpus. The procedures for these tests are discussed above under "Statistical Procedures for Comparing Textual Maps." These tests are divided across two sections. The first section compares Mark to referential narrative. The second section compares Mark to non-referential narrative. The third section brings the results together to produce an overall picture of ideational theme choice and its contribution to understanding genre constraints on narrative texts. Chapter 8 offers some concluding comments and suggestions for further research.

CHAPTER 4: MAPPING MARK ITSELF

General Considerations

Applying the procedure developed in Chapter 3 to Mark is simpler than applying it to the extracanonical literature discussed in later chapters. This ease of application results directly from the fact that the OpenText.org project already annotated Mark, so I had a ready-made set of diagrams with which to work.[1] Given the centrality of Mark and the ease with which I could access the data, I chose to use a full enumeration technique, instead of sampling it. Since I was using the OpenText.org diagrams, I followed their text. The data below also assume that 16:8 is the end of the Gospel.

Troublesome Categorizations

Before I list the tabulated data, a few troublesome categorization decisions call for comment. As noted in the procedure section, a negative particle (οὐ, μή, or occasionally—in conjunction with an aorist subjunctive verb—both) normally does not factor into my analysis. However, on ten occasions, out of a total of 1494 primary clauses, a compound word of which the first element was a negative particle occurred in thematic position.[2] Incidentally, only 2 of these are part of Mark's framework, and both these instances also have another word compounded with the negative particle immediately afterward.[3]

In these cases, I chose to categorize these clause components according to the other element of the compound word on the assumption that the author selected the

[1] These are available online (opentext.org/texts/NT/Mark.html).

[2] Mar.c7_38, Mar.c7_47, Mar.c10_24, Mar.c10_60, Mar.c10_114, Mar.c11_60, Mar.c12_148, Mar.c14_83, Mar.c14_109, Mar.c16_35.

[3] Mar.c12_148 and Mar.c16_35.

compound word to a fill a clause slot because of this part of the word. I should point out that—taking the data as a whole—10 out of 1494 instances is a proportion of .7%, and—even recognizing the skew between narrative and projected discourse among the anomalous instances—this anomaly could only affect the overall picture of projected discourse by 1.07%.[4] Therefore, this decision seems unlikely to have affected my results.

Another anomalous case for which my initial procedure did not account is thematic address. This category occurred 15 times, by definition, as part of projected discourses.[5] My decision in these cases was simply to omit them from analysis. These 15 instances produce a possible error of 2.01%.[6]

In summary, I encountered several hurdles in collecting the data reported in this chapter. None of these were frequent enough to have affected my results meaningfully. My analysis of Mark's register and, thus, genre is thus unhindered.

Quantifying Mark's Thematization

The tables in the next section lay out the values for the observed frequencies and proportions of the realizations of the various semantic choices in the ideational theme system. If Halliday's theory of probabilistic grammar holds, these proportions estimate the grammatical probabilities of the corresponding choices in all Hellenistic Greek texts that match Mark's register, whatever that might happen to be. Before setting out these estimates, however, other descriptions of the data also need to be considered.

---

[4] 8 (anomalies)/748 (total primary clauses of projected discourse) =. 01069.

[5] Mar.c4_155, Mar.c5_131, Mar.c10_56, Mar.c10_70, Mar.c10_93, Mar.c10_145, Mar.c10_203, Mar.c10_207, Mar.c11_89, Mar.c12_56, Mar.c12_85, Mar.c13_3, Mar.c13_203, Mar.c14_151, Mar.c14_160. Note that the frequency picks up as one moves through the Gospel. I am unsure what to make of this pattern.

[6] 15 (anomalies)/748 (total primary clauses of projected discourse) = .02005.

Chapter 3's section on "Statistical Procedures for Quantifying Grammatical Systems" noted that descriptive statistics summarize the shape of the data and, at a bare minimum, one needs to report both the mean and standard deviation, i.e. both the central value and the spread of the data around the central value, to have a complete picture of the data. So far as I have been able to determine, applications of Halliday's model of probabilistic grammar, whether his own or others', have not tended to do this. I have concluded, however, that the variability of the data around the mean value that has typically been reported as an estimate of the grammatical probability potentially also has a significant effect on understanding register variation in a corpus. This will be particularly true in the next two chapters, where I am combining samples from multiple texts together to estimate the grammatical probabilities of the candidate analogies for Mark, but these statistics will still be useful here. First, they will provide a background against which to view observed variations in frequencies within Mark itself; this process forms the basis of "The Shape of Mark's Thematization" below. Second, they also are the background for comparing Mark to the normal distribution in the "Is Mark Normal?" section. Mark is split into 16 bins in each case, corresponding to the chapter divisions. In addition to mean and standard deviation, I also report skewness and kurtosis.

I begin with the least delicate choice in the ideational theme subsystem, choosing whether to place a verb (+Process) or choose some other sort of clause constituent (-Process), or some other sort of clause constituent as the first clause constituent in a primary clause. Mark's framework shows a normalized frequency for +Process of 54.5 selections/100 instances of the system with a standard deviation of 10.37 selections/100 instances of the system. The skewness is -0.41, indicating slightly more data points below the most common value than above it. The kurtosis is 3.47, indicating that slightly more of the variance comes from extreme values than would be ideal. The corresponding data from discourse has a mean of 42.94 selections/100 instances of the system with a standard deviation of 9.43 selections/100 instances of the system. The skewness of these data is 0.14, indicating that more of the data is above the most common value than below it. The kurtosis value is 3.57, indicating that a little more of the variation in observed frequencies between chapters comes from the extreme values than it does for the framework data.

The next subsystem differentiates between the choice to place an adjunct (+Circumstance) or another type of clause constituent typically realized by a noun (-Circumstance) at the beginning of a primary clause. The narrative framework of Mark exhibits a normalized frequency of 55.31 selections per 100 instances of the system with a standard deviation of 18.86 selections per 100 instances of the system. These data exhibit extreme negative skew (-1.43) and a kurtosis of 5.70, indicating a steep peak with much of the data on the edges of the distribution. The data on this subsystem from Mark's projected discourse exhibits a typical value of 33.38 selections per 100 instances of the system with a spread of 17.98 selections per 100 instances of the system. The projected discourse data on this subsystem skews heavily towards the right (2.01) with a very peaked distribution (kurtosis of 7.87).

The third subsystem differentiates between the choice to put a noun functioning as a subject at the beginning of a primary clause (+Subject) and the choice to put some sort of noun functioning as complement, whether direct object, indirect object, or predicate nominative, (collectively, -Subject) first. The data on this subsystem from the narrative framework of Mark exhibit a typical value of 83.38 selections/100 instances of the system for the choice +Subject with a typical spread of 20.88 selections/100 instances of the system.[7] These data skew somewhat to the left (-0.95) and are somewhat flat on top (kurtosis of 2.49). The data on this subsystem from the dialogue portions of Mark give a typical value of 62.68 selections/100 instances of the system with a typical spread of 16.54 selections/100 instances of the system. These data have a long left tail (skewness of -1.09) and a somewhat pronounced peak (kurtosis of 4.10).

The fourth subsystem differentiates between what have traditionally been called direct objects (+Affected), which Hellenistic Greek typically realizes with nouns in the accusative case, and other types of complements (-Affected). The data on this subsystem from the framework of Mark shows a typical value of 30.31 selections/100 instances of the system for +Affected, with a typical spread of 42.41 selections/100 instances of the system.[8] Intuitively, these data have to be positively skewed, because a symmetrical distribution would extend into negative numbers, and the measure of skewness confirms this (0.77). The top of the frequency distribution for +Affected has a flat top (kurtosis of 1.84). The data on this subsystem from the dialogue portions of Mark have a typical value of 69.81 selections/100 instances of the system with a typical spread of 25.69

[7] These data indicate the limitations of studying Mark this way. Notice that the frequency of +Subject in Mark’s framework cannot be above the typical value by a margin of the typical spread, since this would result in a figure of 104.26 selections/100 instances of the system as the upper bound of the region in which 68 percent of the data should fall, which is clearly impossible. This means that the data on subsystem 3 from Mark’s framework *cannot* be normally distributed.

[8] As with the previous subsystem, I can conclusively conclude that these data are not normally distributed, because the typical value and the typical spread produce impossible results. In this case, the impossibility is on the low side, rather than the high side, as was previously the case: subtracting the typical spread of 42.41 selections/100 instances of the system from the typical value of 30.31 selections/100 instances of the system results in -12.1 selections/100 instances of the system as the lower bound of the region in which roughly 68 percent of the data should fall.

selections/100 instances of the system. These data are negatively skewed to a small degree (-0.28) with a flat top (kurtosis of 1.9).

The fifth subsystem differentiates between what have traditionally been called predicate nominatives (+Stative) and what have traditionally been called indirect objects (-Stative), which Hellenistic Greek typically realizes using the dative case. The data on this subsystem taken from the narrative framework of Mark show a typical value of 12.5 selections/100 instances of the system for +Stative, with a typical spread of 46.56 selections/100 instances.[9] As has to be the case with such a low typical value and a high typical spread, these data are radically skewed in a positive direction (2.27) and have a very flat peak (kurtosis of 1.25). The data on this system taken from Mark's projected discourse have a typical frequency of 54.15 selections/100 instances of the system for +Stative with a typical spread of 46.55 selections/100 instances of the system.[10] These data exhibit a small degree of negative skew (-0.19) and a flat peak (kurtosis of 1.25).

Thematization in Mark: The Assembled Data

This section consists of a series of tables. Table 4.1 records the raw frequency counts of the various realizations of the thematic system. As explained in the procedure, only some of the data in Table 4.1 directly represent the frequency of the paradigmatic choices on which a probabilistic approach to grammar is based. Specifically, for the lack of a better way of putting it, these frequency counts represent only the "plus" half of each set of

[9] Once again, I can conclusively conclude that these data are not normally distributed, for the typical value and the typical spread produce impossible results. This case is in the same direction as the previous one, but to a greater degree: subtracting the typical spread of 46.56 selections/100 instances of the system from the typical value of 12.5 selections/100 instances of the system results in -34.06 selections/100 instances of the system as the lower bound of the region in which roughly 68 percent of the data should fall.

[10] Note that this just barely creates an impossible figure for the upper range of the bound in which roughly 68 percent of the data should fall: adding the typical spread of 46.55 to the typical value of 54.19 equals 100.74.

options and the minus option of the most delicate system; one has to reconstruct the frequency of the other “minus” choices from the “plus” halves of more delicate systems. I present the results of this process in two paired sets of tables. Table 4.2 records the raw frequency counts for both halves of each paired opposition within Mark’s narrative framework. Tables 4.3 is the analogous table for projected discourse. In the tables below an asterisk marks the categories with frequencies too low for reliable statistical analysis.

Table 4.1: Overview of Thematization Choice in Mark

	Circumstances	Process	Subject	Complement	Total
Narrative	185	409	139	13	746
Projected	140	322	188	98	748
Total	325	731	327	111	1494

Table 4.2: Frequencies of Theme Choice in Mark’s Narrative Framework

Subsystem	Plus Feature	Minus Feature	Totals
1 (+/- Process)	409 (0.55)	337 (0.45)	746
2 (+/- Circumstance)	185 (0.55)	152 (0.45)	337
3 (+/-Subject)	139 (0.91)	13 (0.09)	152
4 (+/-Affected)	11 (0.85)	2* (0.15)	13
5 (+/-Stative)	1* (0.5)	1* (0.5)	2

Table 4.3: Frequency of Theme Choice in Mark’s Projected Discourse

Subsystem	Plus Feature	Minus Feature	Totals
1 (+/- Process)	322 (0.43)	426 (0.57)	748
2 (+/- Circumstance)	140 (0.33)	286 (0.67)	426
3 (+/- Subject)	188 (0.66)	98 (0.34)	286
4 (+/- Affected)	70 (0.71)	28 (0.29)	98
5 (+/-Stative)	24 (0.86)	4* (0.14)	28

How Reliable Is the Picture of Mark's Thematization System?

The preceding chapter introduced several concepts for evaluating the reliability of quantitative descriptions of grammatical systems like the above tables. First, it suggested that, although Halliday and others have not done so, discussions of grammatical probability could benefit from including a measure of dispersion along with the estimate of the grammatical probability in a text, which is essentially a measure of central tendency (specifically a mean) serving as a summary statistic. Second, it introduced confidence intervals as a succinct notation for describing the degree of uncertainty attached to point estimates like means. Third, it introduced the boxplot as a means for visualizing the reliability of quantitative data. The following tables and charts use these tools to illustrate the reliability of my picture of Mark.

Table 4.4: The Reliability of the Picture of Mark's Framework[11]

Subsystem	95 Percent Confidence Interval
1 (+/- Process)	55±3.57
2 (+/- Circumstance)	55±5.31
3 (+/- Subject)	91±4.54*[12]
4 (+/- Affected)	85±19.41*
5 (+/- Stative)	50±69.30*

Table 4.5: The Reliability of the Picture of Mark's Projected Discourse

Subsystem	95 Percent Confidence Interval
1 (+/- Process)	43±3.55
2 (+/- Circumstance)	33±4.47
3 (+/- Subject)	66±5.49
4 (+/- Affected)	71±8.98*
5 (+/- Stative)	86±12.85*

[11] As a reminder, these numbers result from the procedure for confidence intervals of a proportion, but they are expressed as frequencies for ease of reporting.

[12] The asterisks in these confidence intervals designate the cases where the tests of normality, reported below, indicated that the data's distribution did not conform to the normal distribution sufficiently for parametric tests. Similar assumptions underlie the formula for confidence intervals.

Figure 4.6: Visualizing the Picture of Mark's Framework

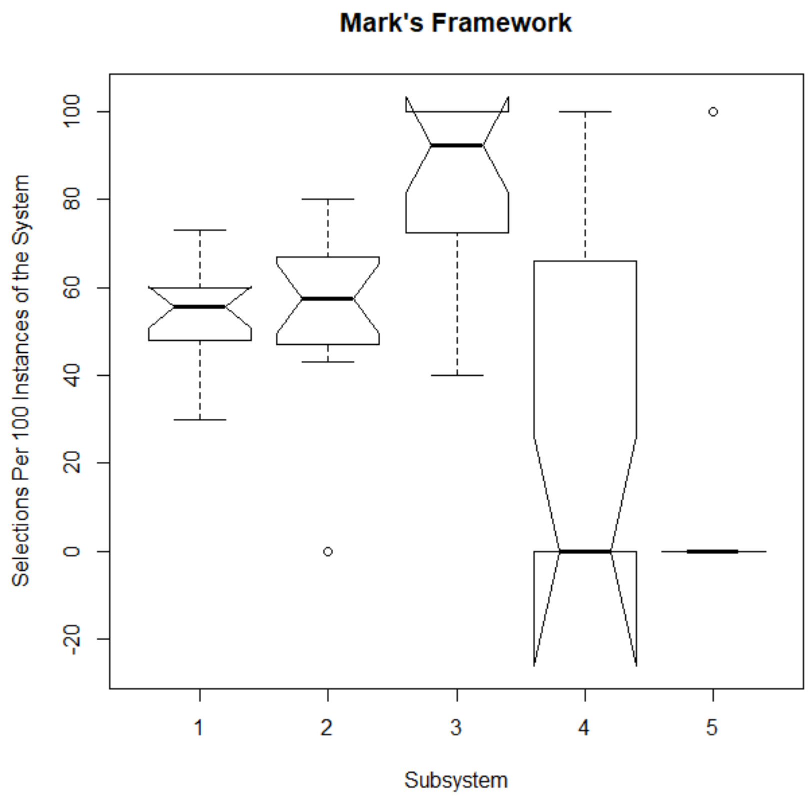
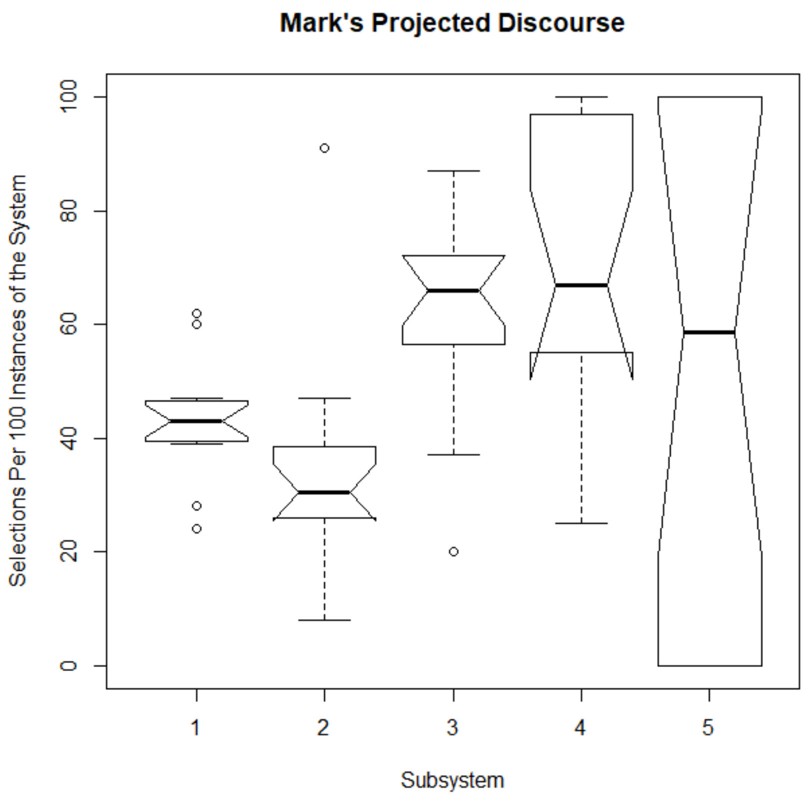


Figure 4.7: Visualizing the Picture of Mark's Projected Discourse



Explanation of the Tables and Figures

Several points emerge from the preceding tables and charts. First, I will briefly mention one while reserving a full discussion of the point for later: the confidence intervals in the table are consistently narrower than the corresponding ones in the boxplots. Since the main difference between the two sets is that I calculated the former from raw frequencies and the latter from normalized frequencies, this points to problems with my normalization procedure.[13] Second, the confidence intervals almost uniformly get wider as one moves from subsystem 1 to subsystem 5, regardless of whether they are calculated

[13] See "Lessons Learned" in Chapter 8.

from raw or normalized frequencies.[14] Both of these points largely have to do with sample size, for the major problem I have found with my normalization procedure is that it groups large numbers of clauses into bins, each of which then has equal weight. This reduced the workload in data collection, but it means that for confidence interval purposes the boxplots have a sample size of sixteen, whereas only two groups of data (subsystems 4 and 5 in the framework) have total frequencies that low. These low-frequency subsystems are also the only ones for which the interval contains impossible values (below 0 or above 100 selections/100 instances of the system). Even with problems in normalization procedure, the number of impossible confidence interval boundaries is also two, though in this case the offenders were subsystem 3 and 4 in the framework. Likewise, the problematic normalization procedure did not prevent fourteen of the twenty confidence interval boundaries in the boxplots from being within their boxes, often by a large margin, whereas several of the problematic boundaries pressed beyond their box by less than 1 selection/100 instances of the system. These data clearly indicate that the data in Tables 4.1 through 4.3 are a reliable picture of Mark's instantiation of the thematization system.

The Shape of Mark's Thematization System: Analysis

I need to answer three related questions about the data presented above before I can compare it with the data for referential and non-referential narratives. First, what do the data indicate about how Mark instantiates the thematization system in his narrative

[14] 90 percent of the confidence intervals (18/20) follow this trend. One of the two that does not is boxplot confidence interval for subsystem 5 in the framework, where the entire box, to say nothing of the confidence interval for the median, sits at 0 selections/100 instances of the system because there are only two instances of +Stative in the framework of Mark. The other instance is more intriguing and will be developed in more detail.

framework? Second, what do the data indicate about how Mark instantiates the thematization system in projected discourse? Third, how do the data relate to the normal distribution? Below are two subsections that answer the first two questions. The third question is more involved, so I will devote an entire section to it.

The Thematization System in Mark's Framework

The task of this subsection is to provide a qualitative interpretation of some patterns discerned in the quantitative distribution of ideational theme choice within Mark's narrative framework.[15] I do not intend this discussion to be exhaustive, only illustrative; no doubt a lengthier subsystem-by-subsystem approach would uncover other intriguing patterns.

The normalized frequencies by chapter for the data on subsystem 1 within Mark's narrative framework differ from the mean by more than the standard deviation only twice, namely chapters 4 and 5. In both cases, these deviations are above the mean, indicating that the deviations are in favor of placing a verb in thematic position over putting something else there. To some degree, this simply points to the content of these chapters: they contain lots of speech and, hence, a lot of speech margins, which typically are primary clauses in the narrative framework, and often speech margins are clauses consisting of simply a conjunction and a verb (e.g. καὶ λέγει, or καὶ ἔλεγεν, of which I count 19 in these two chapters without considering synonyms). Several times in these two chapters, however, an additional clause constituent is inserted into the clause in thematic position in front of the verb. I would not want to make too much of all of these; some seem simply to be the result of logical order, e.g. καὶ εἰσελθὼν λέγει (5:39): Jesus entered

[15] Cf. Halliday, "Linguistic Function," 103.

the synagogue ruler's house before addressing the people inside, instead of yelling at them from the street. On the other hand, some of them seem likely to be significant. In particular, Mar.c5_139 (5:36) contains both a Subject and an Adjunct before the Predicator, and the content of the speech thus introduced is climactic within the pericope: like the unnamed woman whose story is sandwiched within, the synagogue **ruler** must have faith that Jesus can overcome the problem.

Mark's use of subsystem 3 (+/- Subject) in his framework stands out. First, the confidence interval from total frequencies for this group of data is the only one with a reasonable sample size that turned out narrower than the one for the less delicate subsystem on which it depends. While the difference is small (only 0.77 selections/100 instances of the system), it is still striking, given the disparity in instances between subsystems 2 and 3: the formula for standard error, the only variable component of the width of a confidence interval, whether calculated for a mean or a proportion, always has an indicator of variability as the numerator and sample size as the denominator, so a smaller sample having a narrower confidence interval means that the variability decreased more rapidly than the sample size did. Interestingly, this group of data is also the only one where the point estimator of grammatical probability (the mean around which the confidence interval ranges) exceeded 0.9, the cut-off Halliday set for a marked system. This indicates that the choice to place a Complement in thematic position, whether +Affected, +Stative, or -Stative, is a marked choice.

The Thematization System in Mark's Projected Discourse

Subsystem 1 also exhibits significant variation in normalized frequencies within discourse. Four chapters have normalized frequencies that differ from the mean by more then the standard deviation: chapters 1, 7, 8, and 16, with the beginning and end of the text serving as peaks and the midpoint as a valley. This points to a focus on non-activities, on who, when, where, why, and how, rather than on what. Likewise, Mark 7's instantiation of subsystem 2 stands out: 91 selections/100 instances of the system, with nothing else any higher than 47 selections/100 instances of the system. Mark 8's instantiation of subsystem 3 stands out: 20 selections/100 instances of the system, with nothing else less than 37 selections/100 instances of the system. The bulk of thematic complements are concentrated in the middle chapters as well.

Is Mark Normal?

The statistical analyses below and Chapter 7 both involve t-tests. This means I need to determine how well the data from Mark's instantiation of the theme system fit the normal distribution. This section reports the tests by which I make that determination. The first subsection covers the framework, the second subsection covers projected discourse, and the third subsection summarizes the results and their effect on the further progress of the project.

As discussed in chapter 3, I evaluate the data's fit to the normal distribution in several ways. First, I generate a quantile-quantile (Q-Q) plot in R and consider how closely the dots cluster around the reference line. Second, I determine the Pearson correlation coefficient (designated r) between the data sample under consideration and the normalized version R calculated as part of the Q-Q plot. Pearson's r varies between -1 (perfectly inverse relationship) to 1 (perfectly direct relationship), with 0 designating no relationship; for my purposes, I am interested in correlations stronger than r=0.92. Third, I use R to generate a histogram for the data. The histogram helps flush out false correlations where the data look like they are close to the normal distribution but actually have two modes canceling each other out. Next, I report the measures of central tendency—mean, median, and mode—for the sample; these values would be precisely equal if the sample perfectly conformed to the normal distribution. Lastly, I include the skewness and kurtosis of the sample relative to the normal distribution.

Normality of the Framework

The instantiation of subsystem 1 (+/- Process) in Mark's framework conforms to the normal distribution quite well. The QQ plot shows the vast majority of the data clustering quite closely around the reference line, as Figure 1 below graphically illustrates. Only

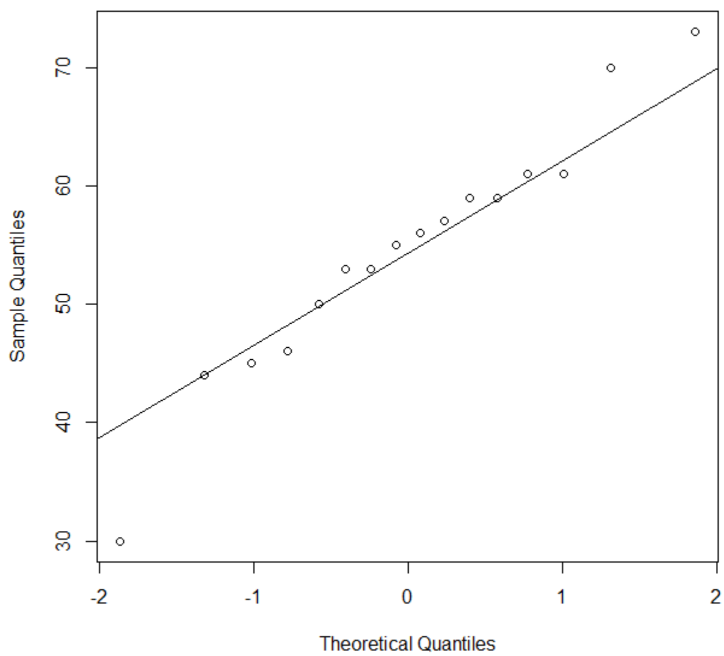


Figure 1: Q-Q plot of subsystem 1 in Mark's framework

three of the data points substantially diverge from the line, whereas four data points actually touch it. In relation to step 2 of the normality testing procedure, the correlation coefficient for this set of data is as high as any in the entire project (r=0.98). The histogram (Figure 2) is far closer to the ideal bell shape than one could reasonably expect a set of 16 observations to be. As a cross-check, the measures of central tendency support this conclusion: the mean is 54.5, the median is 55.5, and the mean of the two values that occur twice—and, as such, are tied for being the mode—is also 55.5. The skewness of this group of data is -0.41, indicating a long left tail. The plot is also slightly more pointy than the normal distribution, with an excess kurtosis of 0.47. Nevertheless, these deviations in shape are small, and thus these data are reasonably consistent with the normal distribution.

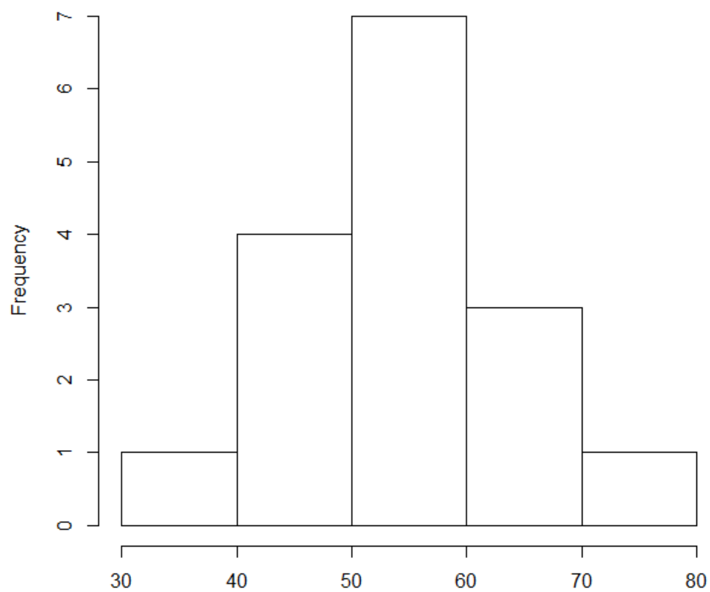


Figure 2: Histogram of +Process in Mark's framework

The applicability of a t-test on the data for subsystem 2 (+/- Circumstance) in Mark's framework is not quite as clear-cut as it was for the previous subsystem. Mark 13 contributes only four clauses to Mark's narrative framework because the majority of the chapter consists of the Olivet discourse. None of the four happens to select

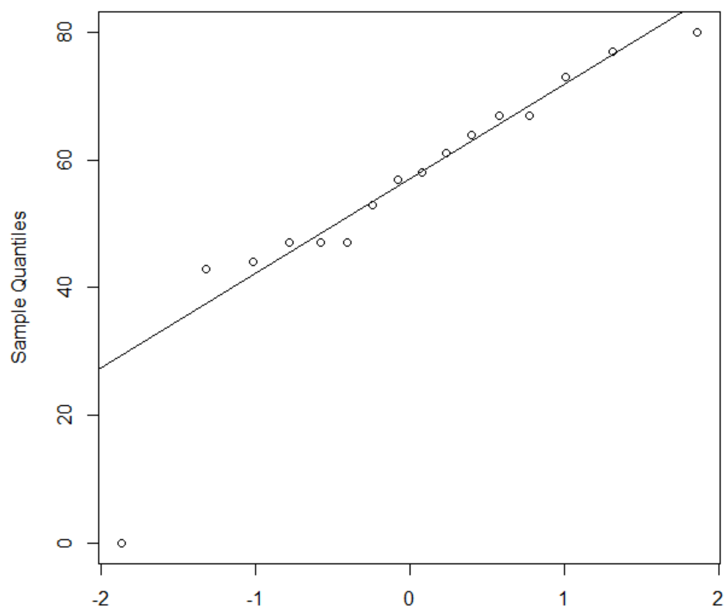


Figure 3: Q-Q plot for subsystem 2 (+/-Circumstance) in Mark's framework

+circumstance, resulting in one instance of 0 selections/100 instances of the system with nothing else below 43 selections/100 instances of the system.[16] As one might expect, the presence of a data point that is far from the rest produces a correlation coefficient lower than for the previous set of data, although it remains just barely within the—admittedly somewhat arbitrary—threshold set for this project (r=0.92). Nevertheless, as Figure 3 shows, the QQ plot still shows discernible clustering. Similarly, the histogram (Figure 4) skews heavily to the right and falls off to quickly to the left to resemble a classic bell-shaped curve. The measures of central tendency also diverge somewhat, ranging from 47 selections/100 instances of the system (the mode) to 58 selections/100 instances of the

[16] See "Lessons Learned" in Chapter 8 for some reflections on cleaning up these data.

system (the median). The skewness for this group of data is -1.43, and the excess kurtosis relative to the normal distribution is 2.70.

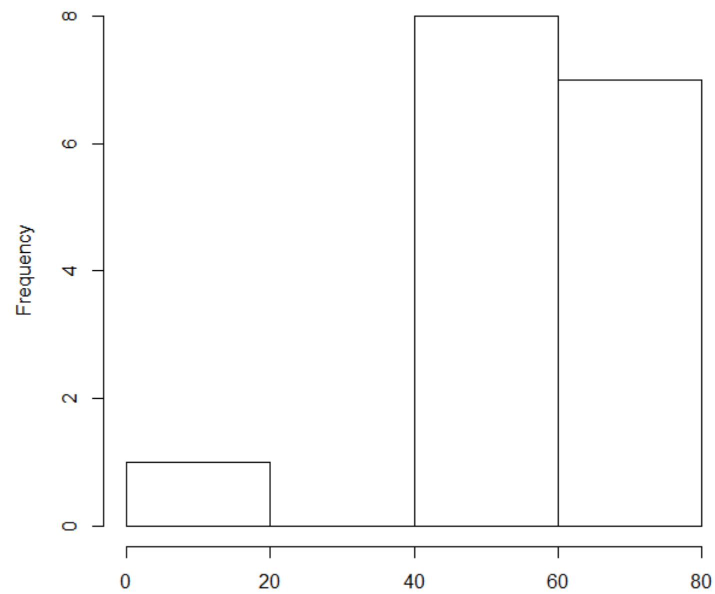


Figure 4: Histogram of +Circumstance in Mark's framework

On the other hand, they are closer to the normal distribution than to any of the other classic distributions (e.g. bimodal or uniform). This fact, the reasonable clustering at the center of the Q-Q plot in Figure 3, and the borderline correlation coefficient combined lead me to believe that a t-test on these data could serve as useful support for the χ^2 analysis that carries the bulk of the statistical weight in this project.

Mark's framework instantiates subsystem 3 (+/-Subject) much differently than the normal distribution would suggest. Comparing Figure 5 with Figures 1 and 3 shows that the data do not cluster around the reference line nearly as well. The histogram (Figure 6) shows that the data actually bear more resemblance to the bimodal distribution than the normal one.

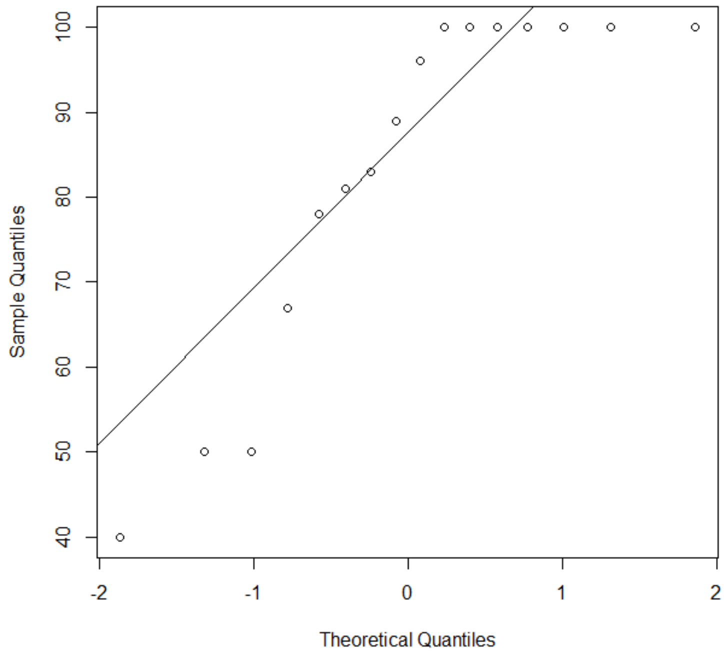


Figure 5:Q-Q plot on subsystem 3 (+/-Subject) in Mark's framework

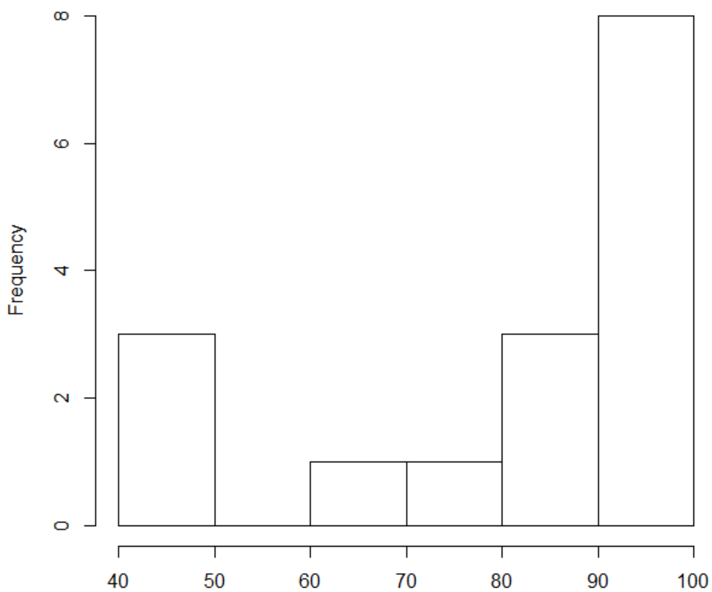


Figure 6: Histogram of +Subject in Mark's framework

In terms of quantitative measures, the measures of central tendency diverge by quite a bit, ranging between 83 instances per 100 selections from the system for the mean to a full 100 instances per 100 selections from the system for the mode. The correlation coefficient indicates poor conformity with the normal distribution as well (r=0.89). The skewness of -0.95 indicates that the data have a long tail on the left side, and the kurtosis relative to the normal distribution of -0.51 indicates that a graph of the data would be fatter than the normal distribution.

The clear skew towards +Subject in Mark's framework means that I will not find much with which to draw a map of subsystems 4 and 5 there, since these susbystems only operate in environments where subsystem 3 selects -Subject. I realized from the outset that this almost guaranteed that the data on these subsystems will not support meaningful χ^2 analysis—to say nothing of t-tests—but I went through the motions in the interest of completeness. As subsystem 4's Q-Q plot (Figure 7) and histogram (Figure 8) show, the extreme values of 0 and 100 selections/100 instances of the subsystem vastly

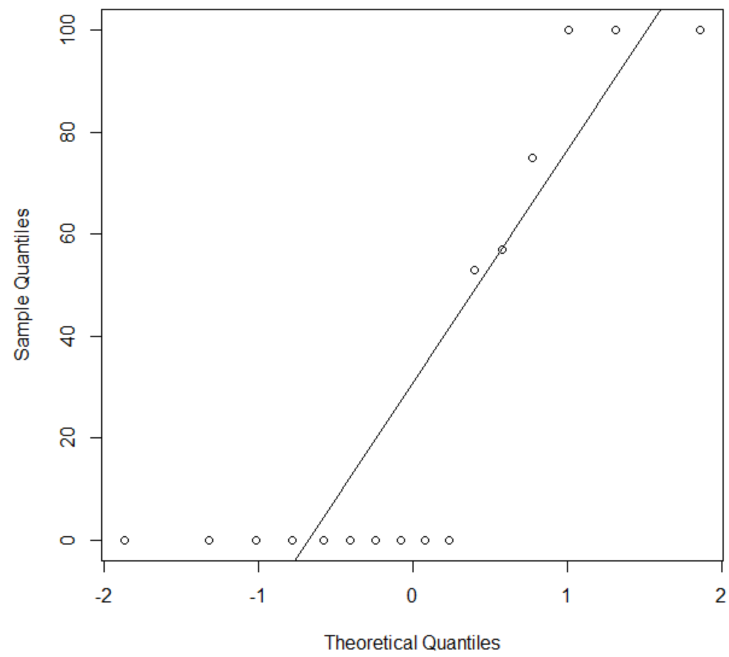


Figure 7: Q-Q plot on subsystem 4 in Mark's framework

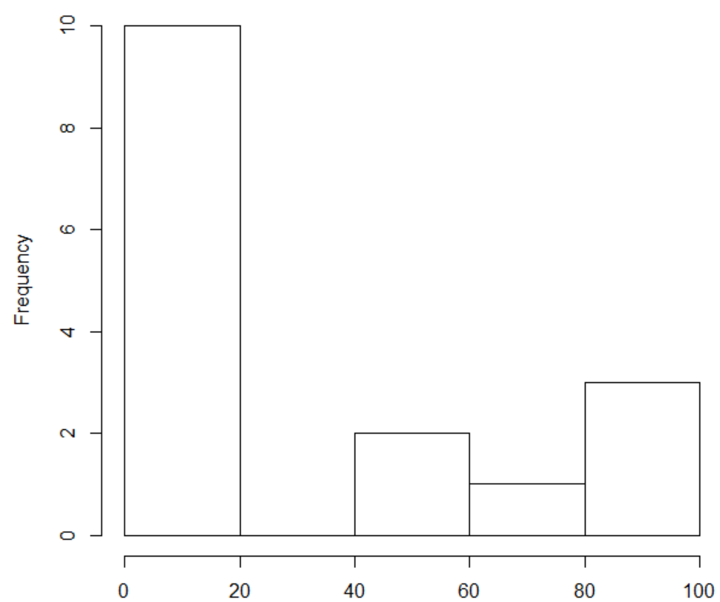


Figure 8: Histogram of +affected in Mark's framework

predominate; only three of the sixteen points fall between these two values. The correlation coefficient falls far short of the threshold (r=0.84). The median and mode are 0 selections/100 instances of the system and 30 selections/100 instances of the system, testifying to the utterly non-normal nature of the data in this group. The skewness of 0.77

indicates that the data has a longer tail than the normal distribution would, but not to a significant degree. Interestingly, the direction of skew is the reverse of that of subsystems 1 through 3. The plot of this data would have a much flatter peak than the normal distribution, as its kurtosis value is 1.16 less than that of the normal distribution. The magnitude of this deviation is significant.

The situation for subsystem 5 (+/- Stative) in the framework is much the same as for subsystem 4, except even more extreme. Note how the reference line for subsystem 5's QQ plot (Figure 9) goes through all the 0 points; the two clauses at the other extreme do not measurably affect it. The correlation coefficient for this group of data is abysmal (r=0.63). As with subsystem 4, the median and mode are both zero, but for subsystem 5 the mean is 12.5. The skewness value of 2.27 and the excess kurtosis value of 3.14 relative to the normal distribution testify to the absolute non-normality of this data. Figure 10 shows the histogram for this subsystem.

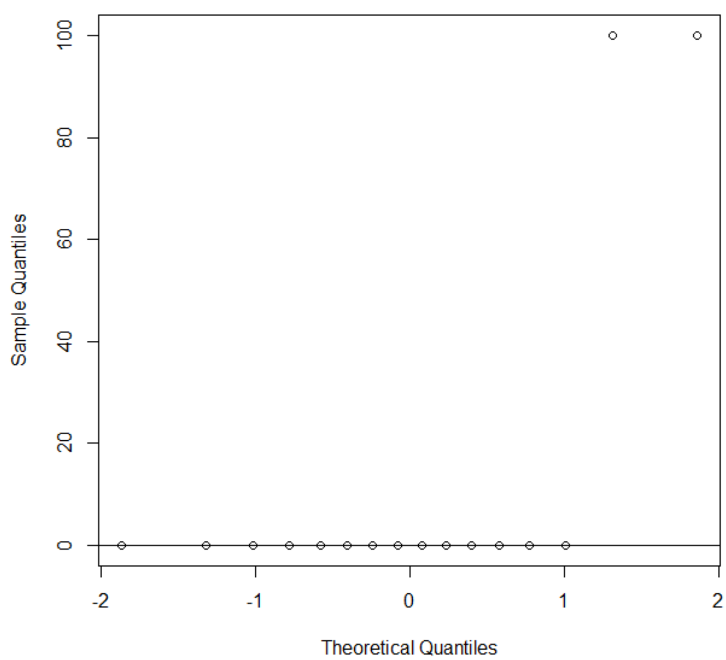


Figure 9: Q-Q plot for subsystem 5 in Mark's framework

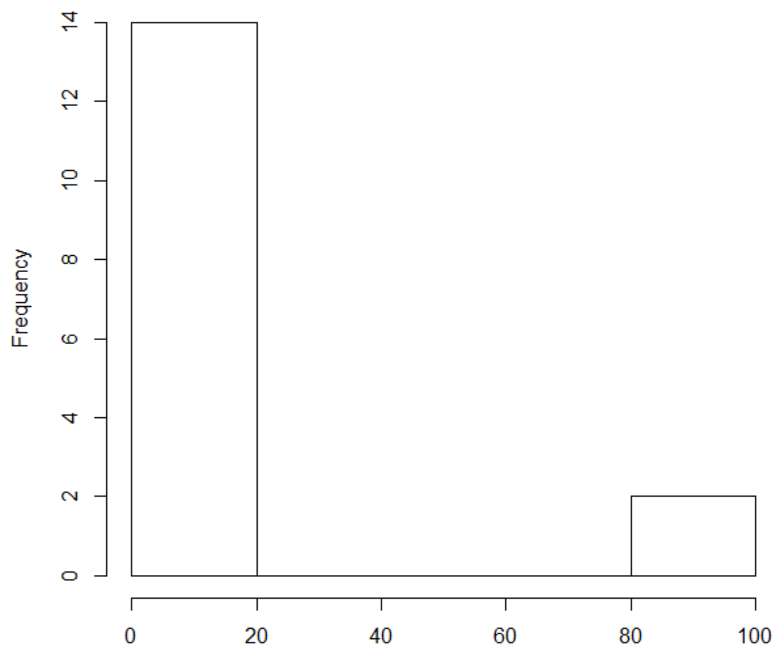


Figure 10: Histogram of +stative in Mark's framework

I recognize that one can sometimes prove that data are non-normal but can never prove that data are normal. Nevertheless, I think the data presented above make a sold case for two sets of data conforming to the normal distribution sufficiently for the use of parametric tests, namely subsystem 1 (+/- Process) and subsystem 2 (+/- Circumstance).

Normality of Projected Discourse

Moving to projected discourse, the data from subsystem 1 (+/- Process) clearly support a t-test. Nearly half the dots in the Q-Q plot (Figure 11) touch the reference line. The quantitative measure of correlation is not stellar, but it is well within the range set for this

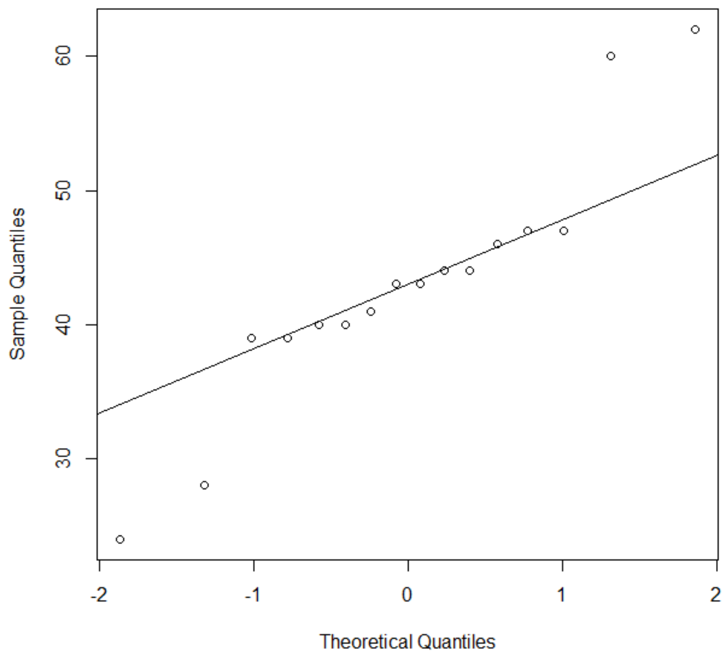


Figure 12: Q-Q plot of subsystem 1 (+/-Process) in Mark's projected discourse

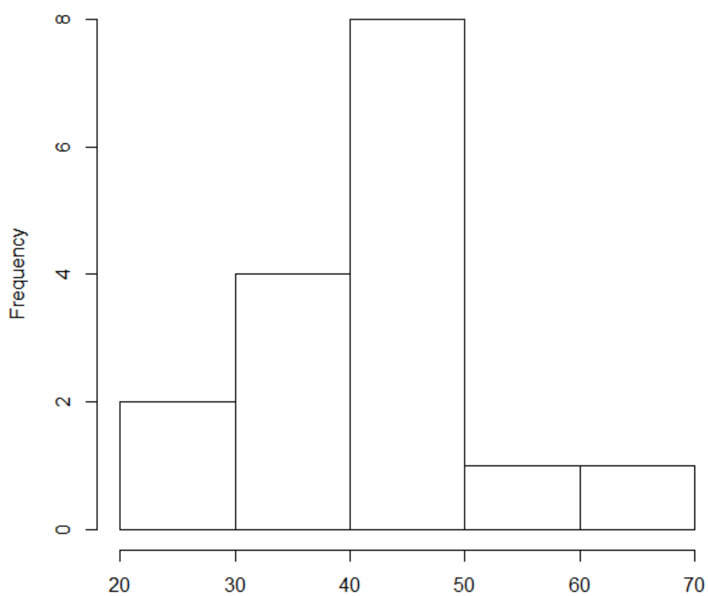


Figure 11: Histogram of +Process in Mark's projected discourse

project (r=0.95). The histogram is quite bell-shaped (Figure 12). The measures of central tendency are quite close: the mean is 43 selections/100 instances of the system, the median is 43 selections/100 instances of the system, and two of the numbers tied for mode are 43 and 44 selections/100 instances of the system. The skewness (0.14) and excess kurtosis (0.57) indicate that the data tail off to the right quicker than they do to the left, but not to an unreasonable degree, and that the shape of the graph is somewhat more peaked than the normal distribution, once again well within reasonable limits.

How suitable t-tests are for analyzing the data derived from Mark's instantiation of subsystem 2 (+/- Circumstance) is somewhat difficult to assess.[17] On the one hand, the quantitative measure of correlation is quite low (r=0.87) and the excess kurtosis relative to the normal distribution is 4.87, which is to say the kurtosis of this group of data is over twice that of the normal distribution. The skewness value of 2.02 is also quite high. On

[17] See "Lessons Learned" in Chapter 8 for some ideas on cleaning up this group of data.

the other, the Q-Q plot (Figure 13) appears to show the data clustering around the

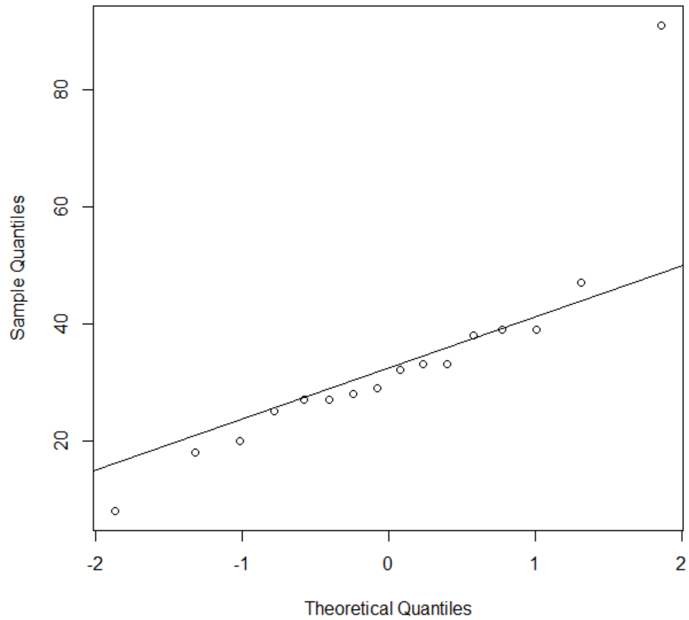


Figure 13: Q-Q plot of subsystem 2 in Mark's projected discourse

reference line.

The measures of central tendency are also fairly close together: the mean is 33 selections/100 instances of the system, the median is between 30 and 31 selections/100 instances of the system, and the mean between the two values tied for mode—27 and 39 selections/100 instances of the system—is 33 selections/100 instances of the system. Likewise, the histogram (Figure 14) has a single peak, which is located near where the measures of central tendency suggest it should be. While I do tend to emphasize

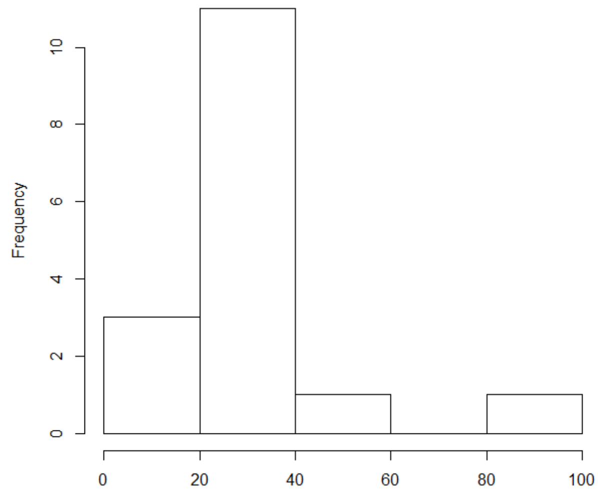


Figure 14: Histogram of +Circumstance in Mark's projected discourse

quantitative measures like the correlation coefficient over qualitative ones, in this case the measures of central tendency, which are also quantitative, agree with my qualitative analysis of the Q-Q plot and histogram. Thus, I decided to include t-tests for this subsystem.

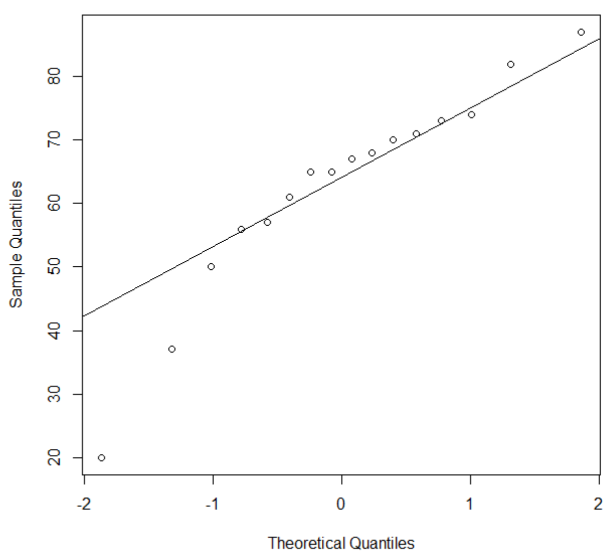


Figure 15: Q-Q plot of subsystem 3 in Mark's projected discourse

Mark's projected discourse instantiates subsystem 3 (+/- Subject) like the normal distribution. The dots on the Q-Q plot (Figure 15) cluster around the reference line.

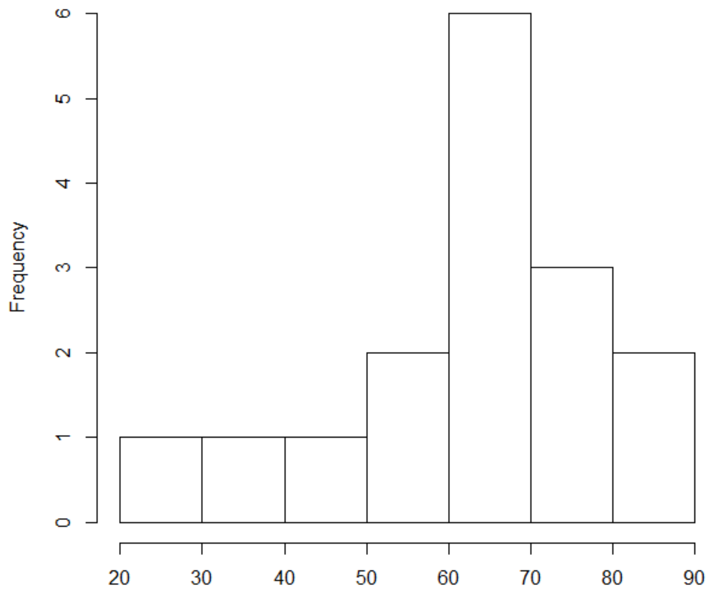


Figure 16: Histogram of +Subject in Mark's projected discourse

The correlation coefficient is well above my stated threshold (r=0.95). The measures of central tendency are quite close to each other: the mean is 63 selections/100 instances of the system, the median is 66 selections/100 instances of the system, and the mode is 65 selections/100 instances of the system. The skewness (-1.1) and excess kurtosis (1.1) are a bit large, but still reasonably consistent with the normal distribution. The histogram (Figure 16) is flatter on the left side than one would prefer, but it still has a single peak located relatively near the center. Thus, I have included t-tests on this subsystem.

Mark's projected discourse contained many more instances of subsystem 4 (+/- affected) than did the framework. Nevertheless, the probabilities associated with the two terms in the subsystem were skew enough that, considered on a chapter-by-chapter basis at least, the sample does not sufficiently conform to the normal distribution to allow for t-tests. The main culprit is the number of chapters where all the instances of subsystem 4 select +affected. These show up as normalized frequencies of 100 selections/100 instances of the system at the top of the Q-Q plot (Figure 17) and the histogram (Figure 18).

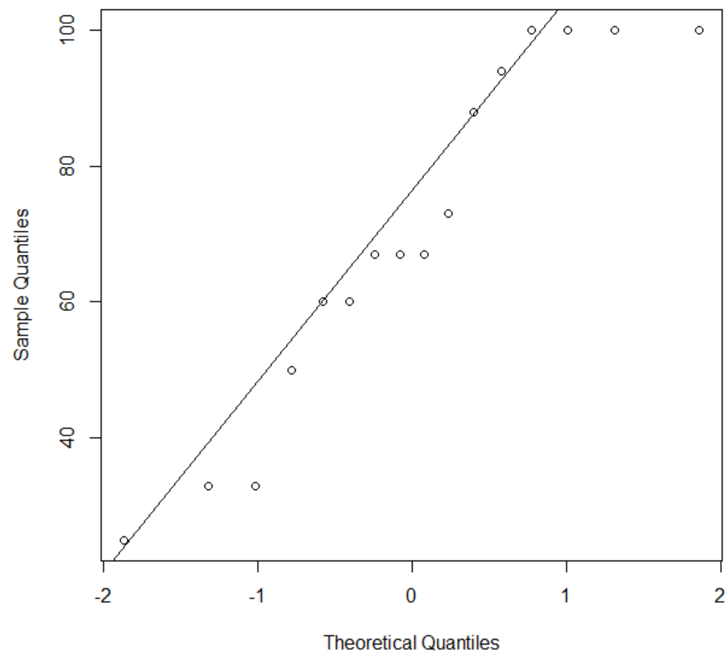


Figure 17: Q-Q plot of Subsystem 4(+/-Affected)

These data are close to being symmetrical, although the left tail is a little longer than the right (the skewness is -0.28). The graph of the data would clearly be broader than the normal distribution with a kurtosis of -1.10 relative to the normal distribution. Figure 18 clearly shows the right side of a bimodal distribution, nearly the reverse of a bell-shaped curve. Consonant with this, the mean (70 selections/100 instances of the system) and median (67 selections/100 instances of the system) are close, while the mode is a full 100 selections/100 instances of the system.

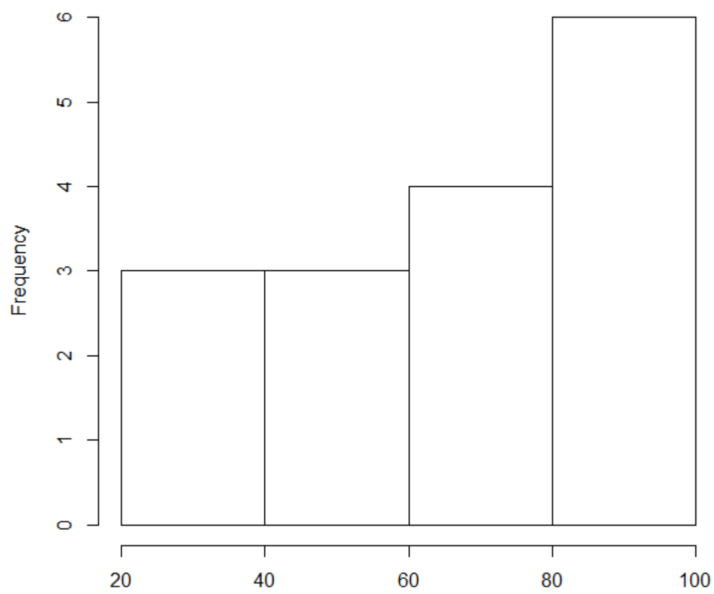


Figure 18: Histogram of +Affected in Mark's projected discourse

The choice -affected and, thus, subsystem 5 (+/-stative) are sufficiently rare that even a sample the size of Mark has only 28 instances. Furthermore, on the occasions where selections from subsystem 5 were made, they tended to cluster, e.g. the four predicate nominatives, realizing +stative, in Mark 14. The normality testing procedure considers the relative frequency with which the terms in a system are chosen, not the absolute frequency, on a chapter-by-chapter basis, thus further diluting the sample. For my normality testing procedure, Mark 14, which selects +stative four times and -stative 0 times, is no different than a chapter that selects +stative once. These factors, as well as the dominance of +stative in the subsystem itself, explain the two sets of values precisely at extreme ends of the spectrum—i.e. where +stative is always or never selected.[18] The

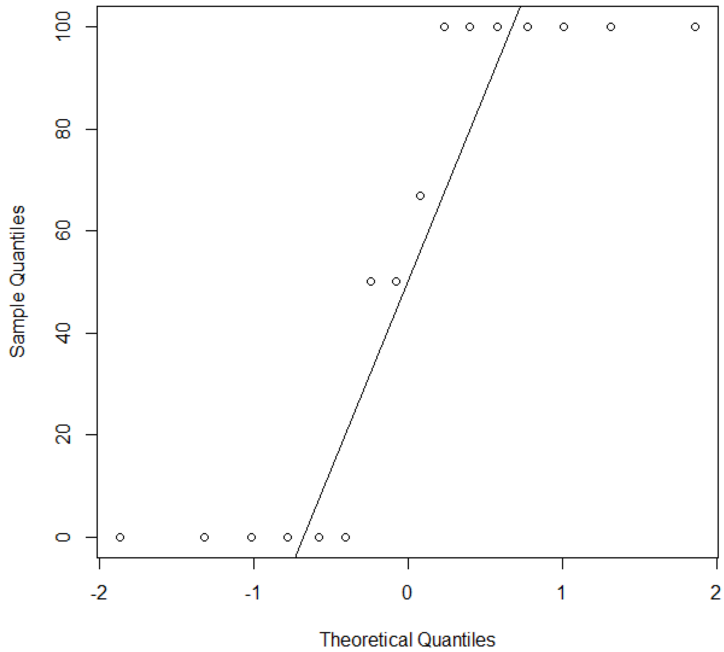


Figure 19: Q-Q plot of subsystem 5 (+/- stative) in Mark's projected discourse

Q-Q plot below (Figure 19) helps illustrate this. In addition, correlation analysis indicates that these data do not allow viable t-tests (r=0.88). The histogram (Figure 20) reinforces this. The skewness is extremely small (-0.19), but the plot has substantially broader shoulders, so to speak, than the normal distribution would, with a kurtosis value of 1.75 less than that of the normal distribution, which is a deviation of over 50 percent. The bimodal nature of the distribution displays itself in the similarity of mean (54 selections/100 instances of the system) and median (between 58 and 59 selections/100

[18] One potential procedural improvement to keep this from happening would be to break the data into smaller units. This would correlate with the desire to choose a smaller basis for normalization.

instances of the system) with a wildly divergent mode (100 selections/100 instances of the system).

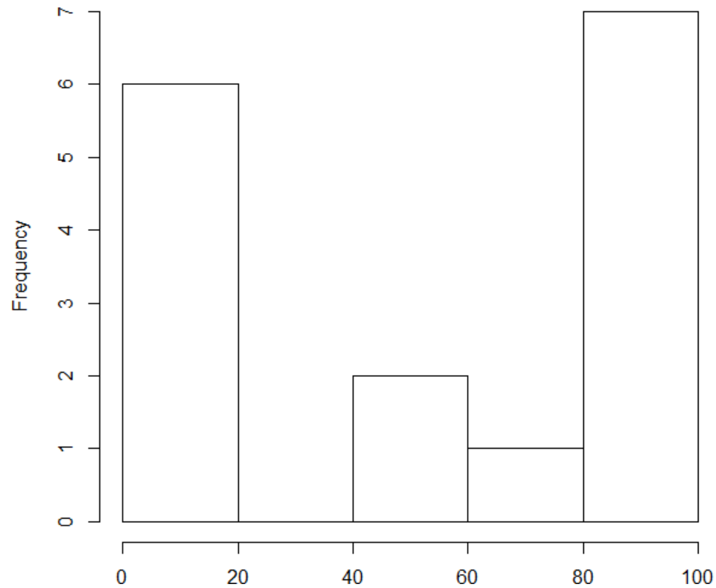


Figure 20: Histogram of +stative in Mark's projected discourse

I recognize that one can sometimes prove that data are non-normal but can never prove that data are normal. Nevertheless, I think the data presented above make a sold case for several sets of data conforming to the normal distribution sufficiently for the use of parametric tests. These include subsystem 1 (+/- Process), subsystem 2 (+/- Circumstance), and subsystem 3 (+/-Subject).

Framework Thematization vs. Discourse Thematization

The preceding two sections show that Mark's narrative framework and projected discourse present distinctly different pictures of the thematization system. This raises the question of whether this system behaves the same way in dialogue as it does in straightforward narration. The data from Mark clearly indicate that they do not, but what are the chances that Mark is an outlier and other texts would show much less deviation? Fortunately, statistical tools provide a means of evaluating this possibility.

From a mathematical perspective this procedure is simply a microcosm of the project as a whole: once again, put in statistical terms, I am comparing multiple sets of numbers obtained as "samples," i.e. representatives, of a larger group, or "population" to which I would like to generalize my results. The samples in the case of the entire project are the constituent narratives—both referential and non-referential—of my comparative corpus; here they are two sorts of linguistic environment within Mark's Gospel. Regardless of the source of the samples, however, a statistical hypothesis test helps adjudicate whether these sets of data are likely to have come from the same theoretical population.

The null hypothesis (H_0) for testing whether I should combine the data from the narrative framework and projected discourse is that the distribution of the various categories (Predicator, Circumstance, etc.) within these environments will be the same. Since I am not concerned about how the two environments differ, if in fact they do, I only need one alternative hypothesis (H_1), namely that the distribution of Predicator, etc. will be different. The test statistics, once again, were the t-test and the χ^2 test. On the assumption that the null hypothesis is true, the grammatical probabilities of the various choices would approximate their frequency in Mark as a whole without differing because they occur in projected discourse or not; thus, the expected frequencies (for the χ^2 test) and the population mean (for the t-test) are set on the basis of the Gospel as a whole.

The t-test comes with the usual caveat regarding the level of the variable and the assumption that the data are normal; here the χ^2 test comes with a caveat as well, namely I have to assume that Mark's choices of theme in the narrative framework immediately surrounding sections of discourse and the choices of theme in the discourse sections do not affect each other. As noted in the tables above, the frequencies of one term in subsystem 4 (-affected) and as a result both terms of subsystem 5, the more delicate subsystem opened up by choosing -affected, are insufficient for reliable statistical analysis. Therefore, three subsections follow, one for each of the subsystems that I can analyze statistically. Then, a final subsection summarizes the results and offers a preliminary explanation for them and how they affect the progress of the project.

Subsystem 1 (+/- Process)

The above discussion of the normality of data taken from Mark supported the normality of both the framework and projected discourse data from subsystem 1, so I performed both a χ^2 test for homogeneity and a t-test. The significance level for both these tests—as with all my hypothesis tests in this project—is 5 percent ($\alpha=0.05$). As the name suggests, the null hypothesis for a χ^2 test of homogeneity is that the samples are homogenous, i.e. that they come from similar populations. If the populations are similar, then the grammatical probabilities associated with the theme system should be similar. In this situation, then, the null hypothesis is that there is no difference between the grammatical probabilities of framework data and projected discourse data. Likewise, the null hypothesis for the t-test is that the framework data and the projected discourse data sprang from populations with similar means. The alternative hypothesis for the χ^2 test is that there is a difference between the grammatical probabilities of the framework data and projected discourse data. The alterative hypothesis for the t-test is that the population means of framework data and projected discourse data differ.

R computed a χ^2 test statistic of 20.264 with 1 degree of freedom. The critical value of χ^2 for 1 degree of freedom and a 5 percent significance level is 3.84. Thus, I can confidently reject the null hypothesis and conclude that Mark's narrative framework and projected discourse data come from different populations.

The null hypothesis for the t-test is that Mark's narrative framework and projected discourse have equal population means; the alternative hypothesis is that the means in question are different. R reported a t-value of 3.30 for 29.73 degrees of freedom. This provides extremely strong evidence for rejecting the null hypothesis, for that t-value would have been significant even if I had chosen .2% as the significance level instead of 5%.

Another interesting point is comparing the skewness and kurtosis of these sets of data. The framework data has a skew of -0.41 and an excess kurtosis of 0.47; the projected discourse data have a skew of 0.14 and an excess kurtosis of 0.57. While the kurtosis of these two sets of data is relatively consistent with one another, the skew differs both in magnitude and direction. These descriptive statistics support the outcome of the hypothesis tests.

Subsystem 2 (+/- Circumstance)

The hypotheses, both null and alternate, for χ^2 analysis are the same here as they were for subsystem 1; only the data for testing it has changed. The total χ^2 value for this subsystem was 29.83—even more significant than the analysis of subsystem 1. This provides extremely strong evidence for rejecting the null hypothesis and concluding that Mark's narrative framework and projected discourse data come from different populations.

As with the χ^2 analysis, the hypotheses for the t-tests here are the same as they were for subsystem 1. R reported a t-value of 3.3672 for 29.73 degrees of freedom. Once again, this provides extremely strong evidence for rejecting the null hypothesis.

The skew and kurtosis once again support concluding that Mark's narrative framework and sections of projected discourse bear witness to different patterns of ideational theme choice. The skew relative to the mean is once again in different directions, though the magnitudes of the skew are much closer in this case than they were for subsystem 1. Once again, both groups of data are more peaked than the normal distribution, though the magnitudes are quite different.

Subsystem 3 (+/- Subject)

The data from this subsystem do not meet the requirements for a t-test. The result of the χ^2 test for subsystem 3 was 35.78. This value also provides strong evidence for rejecting the null hypothesis.[19] The skew and kurtosis come closer to being similar than they did for the first two subsystems. In fact, this is the only time the narrative framework and projected discourse data from Mark skew in the same direction, and the magnitudes are much closer than they are for the other four subsystems as well (-0.95 vs. -1.1). The framework data is somewhat broader than the normal distribution (an excess kurtosis of -0.51), whereas the data from projected discourse is quite a bit more peaked than the normal distribution (an excess kurtosis of 1.1).

Summary

The rest of this study takes for granted that the narrative framework and projected discourse are separate categories. This intuitively seemed the most reasonable course to me, even at the start of the project, despite the temptation to combine them and thus proceed more expeditiously through the material. I reasoned, after all, that I could always

[19] Woods, *Statistics*, 301.

pull them together later, but they would be hopelessly muddled if I started with them together and later decided I should have separated them. This being the case, I undertook the mapping of Mark itself as a useful test case and then performed the statistical analyses recorded here.

These statistical analyses clearly indicated that Mark's framework and projected discourse came from very different populations and, therefore, I should not lump them together—provided the assumptions of the tests had been met. Some of these assumptions I had acknowledged as problems and worked on above (e.g. substantiating which subsystems had approximately normal distributions before performing t-tests on them). One in particular I had not, namely the independence of the data. Once again, it made sense to me that Mark thematized his narrative framework and projected discourse independently because he frequently uses devices like the redundant participle to make clear breaks between the two categories.

Here, again, I was not happy relying solely on intuition, since the hallmark of the contribution I hope to make to the discipline with this study is to use quantitative data where possible. At the time, I could think of no way to test this particular assumption other than letting the proof be in the pudding—that is, do the extra work and see how things play out. Having completed the project, however, I can now return to this point and state that the results of the maps on the comparative corpus seem to bear out this assumption.

CHAPTER 5: MAPPING REFERENTIAL NARRATIVE

This chapter covers the distribution of clause thematization choices among the referential narratives that constitute one of the categories within the comparative corpus discussed in Chapters 2 and 3. This chapter has several sections. First, I need to lay out some background for my discussion of referential narratives. Then, I need to consider the representativeness of the samples. Like the discussion of Mark in Chapter 4, this involves showing that the estimates of grammatical probability derived from the sample are reasonably precise. Unlike the discussion of Mark, however, this chapter involves a second level of sampling, namely the decision to group 1 Esdras and 1 Maccabees into a single group. Initially, I made this decision on qualitative grounds; here, I attempt to show that the data themselves support my intuition**.** Next, as with the discussion of Mark in chapter 4, I will lay out the accumulated data in tabular form, considering the narrative framework and projected discourse separately. In the interests of readability, however, I will split these tables up over two sections, rather than presenting them all at one time, and for similar reasons I will combine the frequencies and proportions into one table.

Background to Discussing Referential Narrative

A variety of issues demand attention before I present the data I have accumulated. Foremost among these are what samples I wound up using from the works discussed in Chapter 3 and delineating the criteria that led me to decide on this sample. Another key consideration is the edition of each text that I used to generate the data. Each document receives a subsection below in which I cover these issues.

First Maccabees

Two troublesome categorization decisions appear here. First, the framework has one example of a compound word involving a negative particle (1Macc.c4_12). Second, the projected discourse within my sample involves one case of thematic address (1Macc.2_177). I handled both cases as I handled their counterparts in Mark: I categorized 1Macc.2_177 as an instance of a thematic Complement on the basis of the rest of the word, and I omitted 1Macc.c4_12 from analysis. This potentially affects the figures in Table 5.3.1 by .2% and the figures in Table 5.4.1 by .8%.

First Esdras

The sample of 1 Esdras involves some troublesome categorizations. First, twice in chapter 2 (1Esd.c2_56 and 1Esd.c2_59), an infinitival clause appears as a thematic Complement. The realization statements of my system network make no allowances for such constructions, forcing some special analysis. After looking at the passage in question, I concluded that these infinitival clauses are more likely functioning like accusative Complements than like nominative or dative ones, so I included them as instances of +affected. With a total of 174 clauses of projected discourse in my sample, these two instances affect my reconstruction of discourse by 1.15%.

Quantifying the Grammatical Systems of Referential Narrative

Having discussed some of the problematic decisions in the annotation process, the next task is to describe the shape of the distribution for the clause thematization system in all of the texts. This description takes the form of reporting the mean, standard deviation, skewness, and kurtosis of each group of data. These descriptive statistics play a part in assessing what statistical tests are appropriate for each group, as described under “Are the

Referential Narratives Normal?" below, and they also serve as the background against which I discuss unusual uses in "The Shape of Referential Narrative's Thematization System" below.

Describing the Shape of the Frameworks of Referential Narrative

Thematic verbs, realizing the choice +Process, dominate the first few chapters of 1 Maccabees. The sample value for the grammatical probability of +Process in this sample is 80 selections/100 instances of the system. The typical spread around this value, as calculated from the normalized values, is 15.75 selections/100 instances of the system. As the reported mean and standard deviation already indicate, these data are negatively skewed (-0.73), for one standard deviation above the mean falls at 95.75 selections/100 instances of the system, barely within the scale, whereas over 30 percent of a normal distribution lies above that point. These data have a flat peak (kurtosis of 2).

Thematic verbs are less prominent in 1 Esdras than in 1 Maccabees. The sample value for the grammatical probability of +Process in this sample is 49 selections/100 instances of the system. The typical spread, calculated from the normalized values, is 9.31 selections/100 instances of the system. This sample is more symmetrical than the one from 1 Maccabees, although it is still somewhat negatively skewed (-.08). This sample is more sharply peaked than the sample from 1 Maccabees as well, but still flatter than the normal distribution (kurtosis of 2.40).

Thematic adjuncts, realizing the choice +Circumstance, are not nearly as common in the sample from 1 Maccabees as are thematic verbs. The sample value for the grammatical probability of +Circumstance in this sample is 32 selections/100 instances of the system. The typical spread, calculated from the normalized values, is 12.34

selections/100 instances of the system. This sample is perfectly symmetrical (skewness of 0). This sample has an extremely flat peak (kurtosis of 1.14).

Thematic adjuncts are proportionately more common in relationship to the noun-based ideational theme choices within the sample from 1 Esdras. The sample value for the grammatical probability of +Circumstance is 42 selections/100 instances of the system. The typical spread around this value, calculated from the normalized frequencies, is 11.10 selections/100 instances of the system. This sample is negatively skewed (-.17). This sample also has an extremely flat peak (kurtosis of 1.71), although not quite as flat as that of the sample from 1 Maccabees.

Thematic subjects, realizing the choice +Subject, are also fairly common in 1 Maccabees. The sample value for the grammatical probability of +Subject is 70 selections/100 instances of the system. The typical spread around that value, calculated from the normalized frequencies, is 13.72 selections/100 instances of the system. This sample is negatively skewed (-0.48). This sample has a flatter peak than the normal distribution (kurtosis of 2.02).

Thematic Subjects are also quite prevalent within 1 Esdras. The sample value for the grammatical probability of +Subject is 84 selections/100 instances of the system. The typical spread around this value, calculated from the normalized frequencies, is 15.34 selections/100 instances of the system. This sample is negatively skewed (-0.41) and has a quite flat peak (kurtosis of 1.32).

Thematic Complements are rare within the sample from 1 Maccabees. Accusative Complements, realizing the choice +Affected, are proportionately much more common than either of the categories involving subsystem 5. The sample value for the grammatical probability of +Affected is 88 selections/100 instances of the system. The typical spread around this value, calculated from the normalized frequencies, is 10 selections/100 instances of the system. This sample is extremely negatively skewed (-1.15). It has a flatter peak than the normal distribution (kurtosis of 2.33).

Thematic Complements are also rare in 1 Esdras. As with 1 Maccabees, accusative Complements are much more common proportionally than either of the categories associated with subsystem 5. The sample value for the grammatical probability of +Affected is 83 selections/100 instances of the system. The typical spread about this value, calculated from the normalized values, is 46.90 selections/100 instances of the system. The normalized frequencies indicate that the sample is quite positively skewed (0.76). The peak is much flatter than the normal distribution (kurtosis of 1.66).

The sample from 1 Maccabees contains only two instances of subsystem 5. As a result, the descriptive statistics are of limited utility. I include them, however, in the interest of completeness. The sample value for the grammatical probability of +Stative is 50 selections/100 instances of the system. The spread calculated from the normalized values is 50 selections/100 instances of the system. This sample is very positively skewed (1.15). It has a flatter peak than the normal distribution (kurtosis of 2.33).

The sample from 1 Esdras also contains two instances of subsystem 5, though in this case both instances are +Stative. The standard deviation calculated from the normalized frequencies is 48.80 selections/100 instances of the system. This sample is positively skewed (0.95). This sample has a much flatter peak than the normal distribution (kurtosis of 1.9).

Describing the Shape of Projected Discourse in Referential Narrative

Thematic verbs, realizing the choice +Process, are quite common in the sample from 1 Maccabees, though not as common as in the framework. The sample value for the grammatical probability of +Process is 51 selections/100 instances of the system. The typical spread around that value, calculated from the normalized frequencies, is 27.98 selections/100 instances of the system. This sample is somewhat positively skewed (0.25). This sample has a quite flat peak relative to the normal distribution (1.37).

Thematic verbs are also common in projected discourse within the sample from 1 Esdras. The sample value for the grammatical probability of +Process is 34 selections/100 instances of the system. The typical spread around this value is around 14.01 selections/100 instances of the system. This sample is more positively skewed than the corresponding sample from 1 Maccabees (0.74). This sample has a flat peak relative to the normal distribution (2.07).

Thematic adjuncts, realizing the choice +Circumstance, are not particularly common in projected discourse within the sample from 1 Maccabees. The sample value for the grammatical probability of +Circumstance is 30 selections/100 instances of the system. The typical spread around this value, calculated from the normalized frequencies, 9.54 selections/100 instances of the system. The sample is slightly positively skewed (0.19). This sample has a far flatter peak than the normal distribution (kurtosis of 1.5).

Thematic adjuncts are about as proportionately common in projected discourse within the sample from 1 Esdras. The sample value for the grammatical probability of +Circumstance is 31 selections/100 instances of the system. The typical spread around this value, calculated from the normalized frequencies, is 15.11 selections/100 instances of the system. This sample is very positively skewed (1.50). This sample has a steeper peak than the normal distribution (kurtosis of 3.69).

Thematic subjects, realizing the choice +Subject, are the most common category after thematic verbs in the projected discourse of the sample from 1 Maccabees. The sample value for the grammatical probability of +Subject is 71 selections/100 instances of the system. The typical spread around this value is 29.19 selection/100 instances of the system. This sample is quite negatively skewed (-0.68). This sample has a far flatter peak than the normal distribution (kurtosis of 1.5).

Thematic subjects are also the most common category after thematic verbs in the projected discourse portions of the sample from 1 Esdras. The sample value for the grammatical probability of +Subject is 54 selections/100 instances of the system. The typical spread around this value calculated from the normalized frequencies is 22.18 selections/100 instances of the system. This sample is moderately skewed in a negative direction (-0.28). This sample has a flatter peak than the normal distribution (kurtosis of 2.17).

Thematic complements of all stripes are relatively infrequent in the projected discourse portions of 1 Maccabees, but accusative Complements, realizing the choice +Affected, are more common than either of the two categories associated with subsystem 5. The sample value for the grammatical probability of +Affected is 57 selections/100 instances of the system. The typical spread around this value, calculated from the normalized frequencies, is 9.81 selections/100 instances of the system. This sample is quite positively skewed (0.71). This sample has a much flatter peak than the normal distribution (kurtosis of 1.5).

Thematic Complements are infrequent in the projected discourse portions of the sample from 1 Esdras, though apparently less so than in the sample from 1 Maccabees. Accusative Complements, realizing the choice +Affected, are once again more common than either of the categories associated with subsystem 5. The estimate derived from this sample for the grammatical probability of +Affected is 83 selections/100 instances of the system. The typical spread, calculated from the normalized frequencies, around this central value is 11.64 selections/100 instances of the system. This sample skews heavily in the positive direction, though not to quite the same degree as the sample from 1 Maccabees. The sample has a much flatter peak than the normal distribution, though not quite as flat as that of the sample from 1 Maccabees (kurtosis of 1.87).

The sample from 1 Maccabees contains only three instances of subsystem 5, all of which chose +Stative. The estimate derived from this sample for the grammatical probability of +Stative is thus 100 selections/100 instances of the system. The normalized frequencies for +Stative are divided between cases of 0 selections/100 instances of the system and 100 selections/100 instances of the system, so the standard deviation calculated from these values is 57.74 selections/100 instances of the system. This sample registers as having no skew because the skewness function is using the normalized frequencies with no information about the measures of central tendency, meaning it cannot distinguish between a normal distribution with a mean of 50 selections/100 instances of the system and a case as extreme as this one. Likewise, the sample registers as having an extremely flat peak (kurtosis of 1), but this is likely due to the lack of variation in the sample.

Instances of subsystem 5 are more common in the projected discourse portions of the sample from 1 Esdras than in 1 Maccabees, but they are still very infrequent relative to the other categories. The estimate derived from this sample for the grammatical probability of +Stative is 38 selections/100 instances of the system. The standard deviation calculated from the normalized frequencies is 44.49 selections/100 instances of the system. The sample appears to be somewhat positively skewed (0.27). The sample appears to have a much flatter peak than the normal distribution (kurtosis of 1.49).

The Validity of "Referential Narrative" as a Genre Category

This section adds quantitative support to the qualitative assignment of 1 Maccabees and 1 Esdras to the same genre category, what I am terming "referential narrative." This quantitative support involves three tools: (1) the descriptive statistics reported in the preceding section, (2) confidence intervals for the relative frequencies on which the sample means (i.e. estimates of grammatical probability) are based, and (3) boxplots with confidence intervals for the median of the normalized frequencies. Tables 5.1 and 5.2 contain the confidence intervals of the relative frequencies for the framework and projected discourse, respectively; the boxplots appear as part of the explanations that follow.

Table 5.1: Confidence Intervals of the Relative Frequencies (Framework)

Subsystem	1 Maccabees	1 Esdras
1 (+/-Process)	80±3.94	49±6.09
2 (+/-Circumstance)	32±10.35	42±8.39
3 (+/-Subject)	70±12.34	84±8.19
4 (+/-Affected)	88±21.56	83±19.68
5 (+/- Stative)	50±69.30	100±100

Table 5.2: Confidence Intervals of the Relative Frequencies (Discourse)

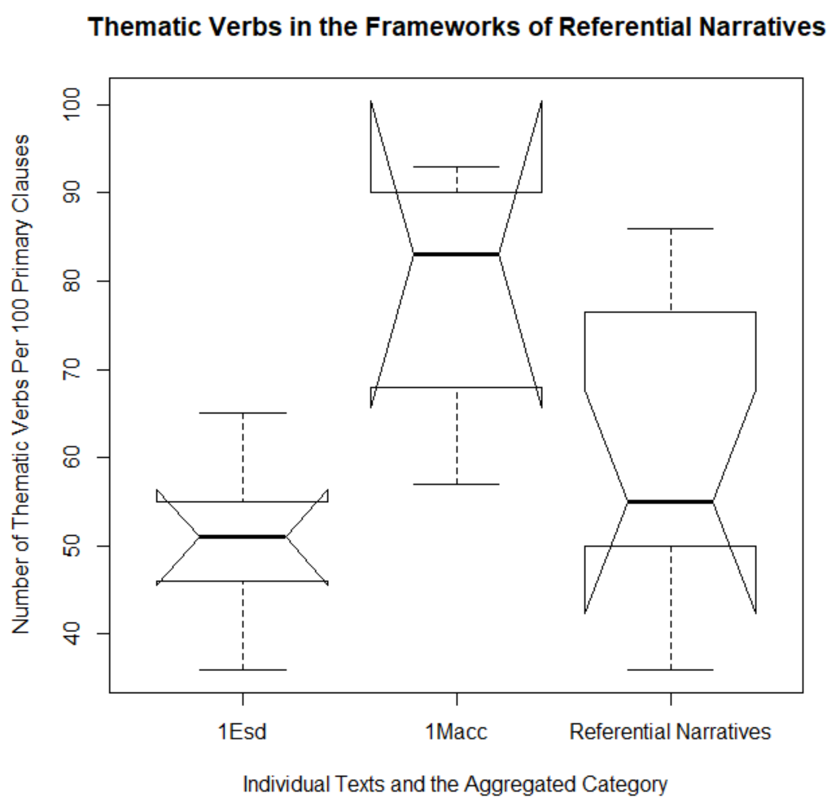
Subsystem	1 Maccabees	1 Esdras
1 (+/-Process)	51±9.10	34±7.04
2 (+/-Circumstance)	30±11.90	31±8.49
3 (+/-Subject)	71±13.89	54±12.13
4 (+/-Affected)	57±36.68	78±13.53
5 (+/- Stative)	100±100	38±33.64

Do the Frameworks Point to a Coherent Category?

Overall, the data on the use of subsystem 1 (+/- Process) in the framework does not support the usefulness of combining these two texts into a single genre category. While both samples are negatively skewed and flatter on top than the normal distribution, the degree of skew in the sample from 1 Maccabees is much greater. Likewise, the overall means and the spread around this mean calculated from the normalized values are also quite different. Comparing the confidence intervals in the first row of Table 5.1 shows a gap of over 20 selections/100 instances of the system between the lower bound of the confidence interval for 1 Maccabees and the upper bound of the confidence interval 1 Esdras.

The boxplot below (Figure 5.1) paints the proverbial thousand words of support for the above consideration of descriptive statistics and the confidence intervals for the relative frequencies of +Process. First, the plot clearly shows that the confidence intervals for the median of the normalized frequencies do not overlap. This is particularly striking given how wide the confidence intervals are in the first place. By their nature, wide confidence intervals are likely to overstate the amount of overlap between samples, so having a gap of about 9 selections/100 instances remaining means that these two samples are unlikely to have come from the same population, assuming that each sample is in fact representative of the population from which it came.

Regarding this last question of sample representativeness, all four confidence interval boundaries for the two component texts have triangles protruding from their respective boxes.[1] This indicates that the normalized frequencies do not provide reliable pictures of the population(s) from which the samples sprang. If the sample median does not reliably estimate the population value, the confidence interval for that median is not a reliable guide for ascertaining whether the two samples belong to the same population.



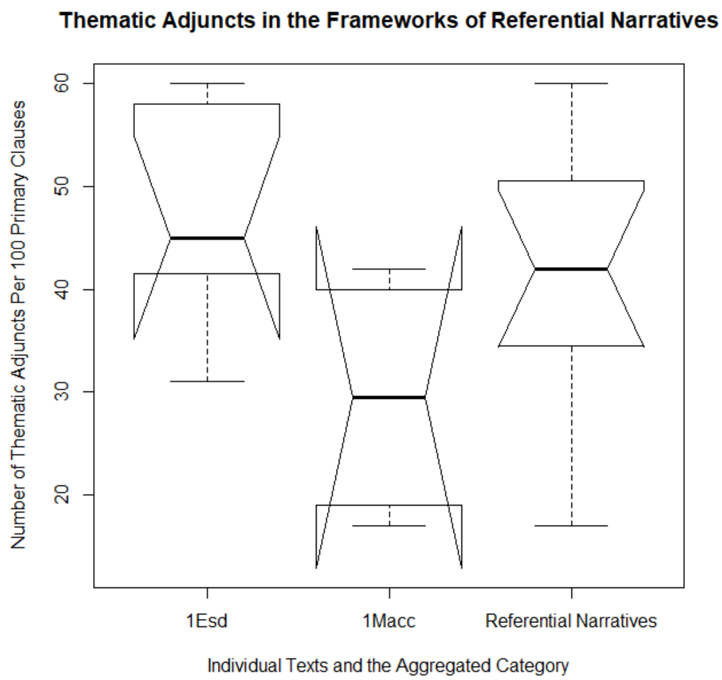
In other words, as they currently stand, the confidence intervals in the figure above count neither for nor against considering 1 Esdras and 1 Maccabees a single group because it is impossible to tell from the two samples whether there are two populations

[1] The triangles at the bottom end of the box for the sample from 1 Esdras are difficult to see in the figure because I reduced the scale to what would fit on the page. Inspecting the numerical output, however, demonstrates that the confidence intervals protrude beyond the box, albeit by less than 1 selection/100 instances of the system.

represented or a dispersed one with a median somewhere in between the medians of the two samples. The way to resolve this difficulty, as with all problematic confidence intervals, is to collect more data. There are two sorts of relevant data in this case. First, lengthening the samples already collected will almost certainly narrow the respective confidence intervals, and the sample median will shift as well, unless the new data are precisely symmetric around the median of the current data. If the medians shift substantially closer, their confidence intervals may overlap, despite the intervals themselves having narrowed; if the medians shift further apart while the confidence intervals narrow enough to be reliable, this would constitute firm evidence that I should not combine 1 Esdras and 1 Maccabees into a single genre category. Second, taking samples from additional documents would also help. Medians for the additional samples clustering in between those of the two current samples would support this being one, wide distribution; medians for the additional samples that did not so cluster would point to differing distributions, potentially also clarifying which of the current samples is the odd one out.

Summing up the preceding discussion, the data on subsystem 1 (+/- Process) from the framework do not support treating 1 Esdras and 1 Maccabees as a single genre category. The basic descriptive statistics are quite varied and the confidence intervals, both those calculated from raw counts and normalized frequencies, do not overlap. Given the concerns about sample representativeness raised by the boxplot, however, additional data, in the form of either lengthening the current samples or adding samples from other documents, could potentially remedy this.

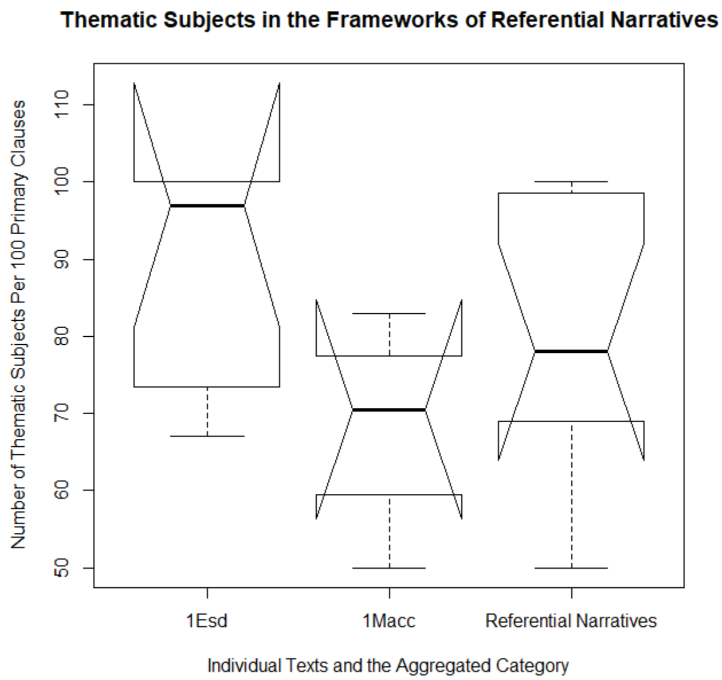
The data on subsystem 2 (+/- Circumstance) from the framework strongly support treating 1 Esdras and 1 Maccabees as a single genre category. The descriptive statistics are quite similar: the standard deviations calculated from the normalized frequencies differ by only roughly 1.25 selections/100 instances of the system, both samples have peaks much flatter than the normal distribution (the kurtosis of both samples (the kurtosis of both samples is below 2), while the skewness is only marginally different (-0.17 vs. 0). Moreover, the confidence intervals calculated from the relative frequencies not only overlap, but the mean of the wider confidence interval falls less than 2 selections/100 instances of the system outside the narrower confidence interval, indicating that the degree of overlap is substantial. The boxplot for these data is below.



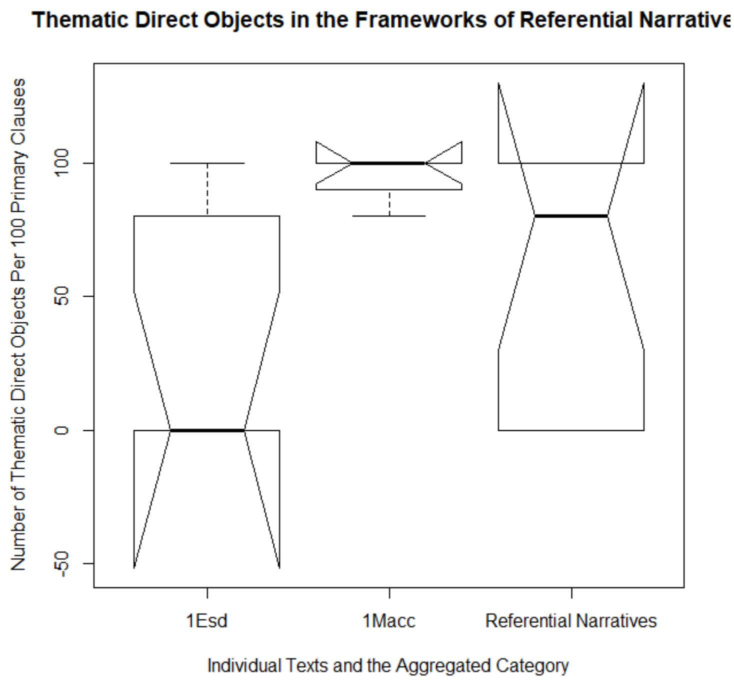
This boxplot offers sheds mixed light on whether 1 Esdras and 1 Maccabees belong to the same genre. On the negative side, both samples have problematically wide confidence intervals, marked in the figure by protruding triangles. Three of the four boundaries are wider than their respective boxes, and the confidence interval for the median normalized frequency in the sample from 1 Maccabees takes up a full third of the theoretically possible values. On the positive side, the confidence intervals overlap substantially, although, granted, that means less than it otherwise would in light of the width of the intervals and the questionable generalizability of the samples. Furthermore, in contrast to the previous boxplot, the boxplot of the aggregated data looks ideal. Two inferences follow from this observation. First, if the two samples represented two markedly different distributions, surely the boxplot of the aggregated data would be far messier, as seen with the preceding subsystem. Second, the one distribution would seem to have a population median above 40 selections/100 instances of the system, which in turn implies that lengthening the samples should pull their respective medians closer to one another as the sample becomes a more accurate estimator of the population value. While lengthening the samples to confirm this, instead of merely pointing out that statistical principles indicate it should be the case, would be preferable, I think it is reasonable to treat the boxplot as providing limited support for combining the two groups, especially given that all the other indicators also pointed in that direction quite strongly.

Turning to subsystem 3 (+/- Subject), the more reliable tools, descriptive statistics and the confidence intervals calculated from the relative frequencies, are consistent with treating 1 Maccabees and 1 Esdras as a single genre category. The standard deviations calculated from the normalized frequencies for the respective samples differ by less than 2 selections/100 instances of the system. Both samples are negatively skewed to almost precisely the same degree (-0.48 and -0.41). Both samples have a flatter peak than the normal distribution, although the peak of the sample from 1 Esdras (kurtosis of 1.32) is quite a bit more so than that of the sample from 1 Maccabees (kurtosis of 2.02). Likewise, the confidence intervals calculated from the relative frequencies overlap, though not to the degree that the respective means are within the confidence interval for the other.

The boxplot for this group of data is more problematic. Given the problems I have observed with my normalization procedure over the course of the study, at least some of the problems seem to derive from the boxplot's reliance on normalized frequencies. Regardless of the cause, however, the boxplot has extremely wide confidence intervals, which in this case of the sample from 1 Esdras extend well into the range of impossible values (the upper bound is 113 selections/100 instances of the system), and three of the four boundaries have protruding triangles indicating that the sample is a poor estimate for the population. Moreover, unlike the corresponding plot for subsystem 2, simply aggregating the data does not produce a coherent plot. The confidence intervals for the sample medians overlap by roughly 3.5 selections/100 instances of the system, which is almost meaningless given the widths of the intervals. These factors, among others, suggest that additional data is needed before the confidence intervals for the boxes will clearly indicate whether the two samples belong to a single group. As discussed with subsystem 1, this additional data could take the form of collecting additional data from these two documents, collecting data from additional documents, or both. Lengthening the sample from 1 Esdras seems particularly valuable in this case because the confidence interval currently includes such a range of values I know will not be replicated in the additional data (because they are impossible). This means that the median from the sample from 1 Esdras is likely to come down, though whether it will do so sufficiently for a narrower confidence interval to overlap remains unclear. In any case, the two other indicators pointed to placing 1 Esdras and 1 Maccabees in the same group, and a problematic boxplot is not sufficient to overturn them.

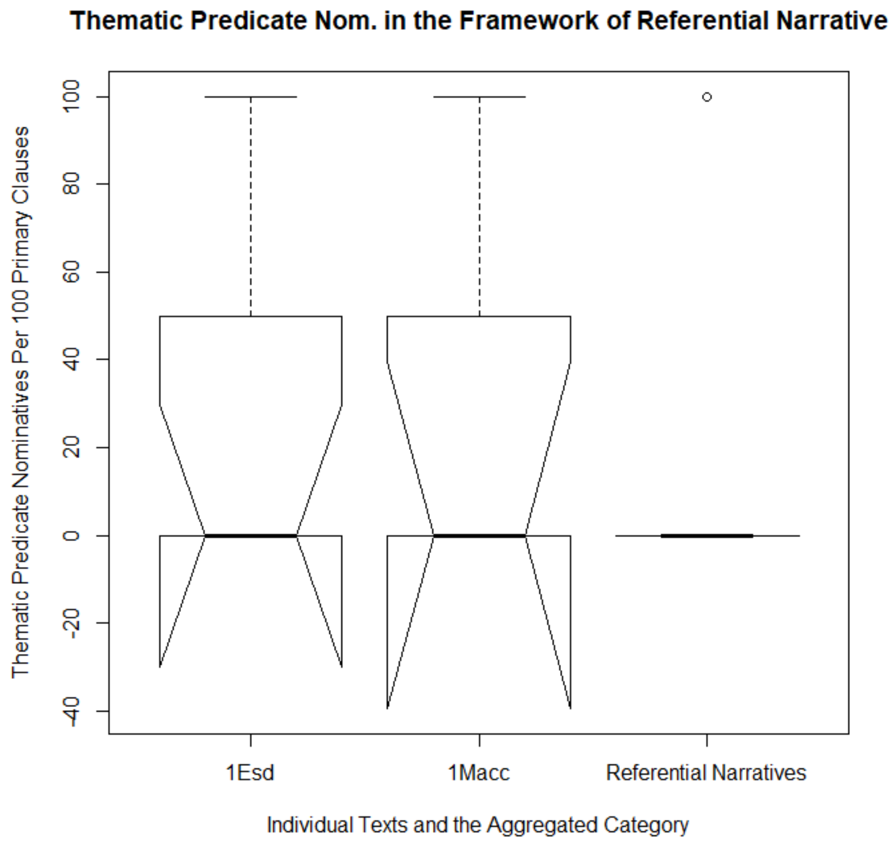


The number of thematic Complements in the framework of these two samples is small enough that the tools for determining whether 1 Esdras and 1 Maccabees belong to the same genre category may well yield unreliable results, for sample size is a major component of the formulas involved. Nevertheless, for the sake of completeness and comparison, I am including subsystems 4 and 5 in this discussion. For subsystem 4, which distinguishes between direct objects (+Affected) and other forms of Complements (-Affected), only kurtosis points to the samples being similar, with both samples having a flatter peak than the normal distribution: the standard deviations are wildly divergent (a gap of over 30 selections/100 instances of the system) and, while both samples are skewed, the skew is in opposite directions. Similarly, although the confidence intervals derived from the raw frequencies overlap, they both include values over 100 selections/100 instances of the system, so the amount of stock one can place in them is limited. The boxplot based on the normalized frequencies is below.



The boxes in this diagram are highly problematic. The confidence intervals for all three boxes extend into *a priori* impossible territory: the lower bound of the confidence interval for 1 Esdras's sample median is -51 selections/100 instances of the system, while the upper bounds for the other two confidence intervals both exceed 100 selections/100 instances of the system (roughly 108 and 130 selections/100 instances of the system, respectively). The problems I have already observed with the normalization procedure are to blame for much of this. For example, the median of the sample from 1 Esdras sits at 0 selections/100 instances of the system, a value that guarantees that a symmetrical confidence interval will dip into impossible values, because over half the chapters in the sample have no thematic Complements.

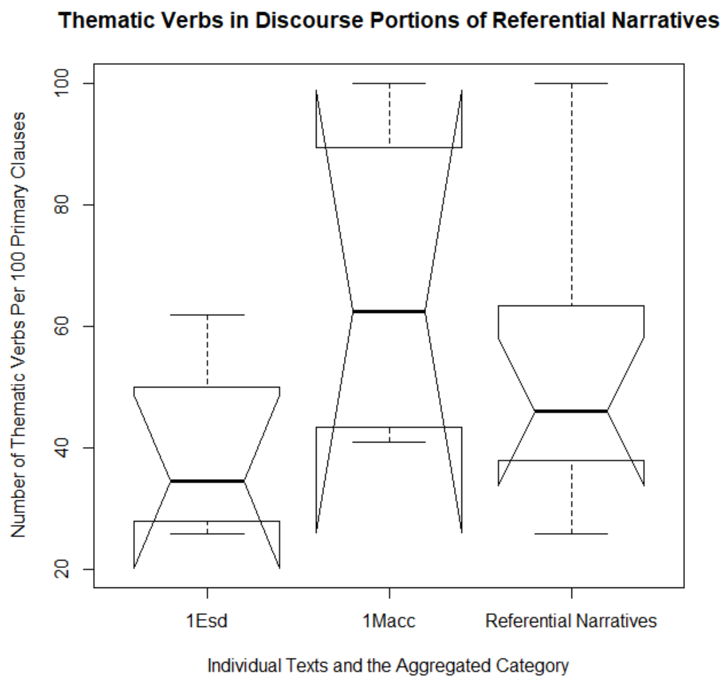
The data on subsystem 5 (+/- Stative) is even more limited than for subsystem 4 (+/- Affected), meaning that the usefulness of these data for determining whether 1 Esdras and 1 Maccabees belong to a coherent group is also more limited. Even so, the descriptive statistics match fairly well, even if only to indicate both samples are equally bad representations of their respective population. The standard deviations calculated from the normalized frequencies are 48.80 selections/100 instances of the system and 50.00 selections/100 instances of the system, which is a difference of 1.2 selections/100 instances of the system, but their size in absolute terms essentially indicates that any theoretically possible value is equally likely. Both samples are extremely positively skewed and substantially flatter on top than the normal distribution as well. The confidence intervals based on the relative frequencies illustrate the import of standard deviations around 50 selections/100 instances of the system: the confidence intervals overlap by default because any theoretically possible value (and some impossible ones, for that matter) falls within both confidence intervals. The confidence intervals for the sample median in the boxplot based on the normalized frequencies paint an even starker picture.



On the positive side, the top halves of the plot look almost identical, with both notches well inside the box. On the negative side, both confidence intervals extend well below the theoretical minimum. This is because over half of the datapoints in both categories are at 0 selections/100 instances of the system. Though the confidence intervals overlap, here again it is because both of them are a poor estimate for any population that could exist. After all, a population median of 0 selections/100 instances of the system would mean that the system did not exist: any value above 0 selections/100 instances of the system, i.e. any instances of +Stative, would have to be balanced by a section where +Stative was selected a negative number of times, which is flatly impossible.

Do the Dialogue Portions Point to a Coherent Category?

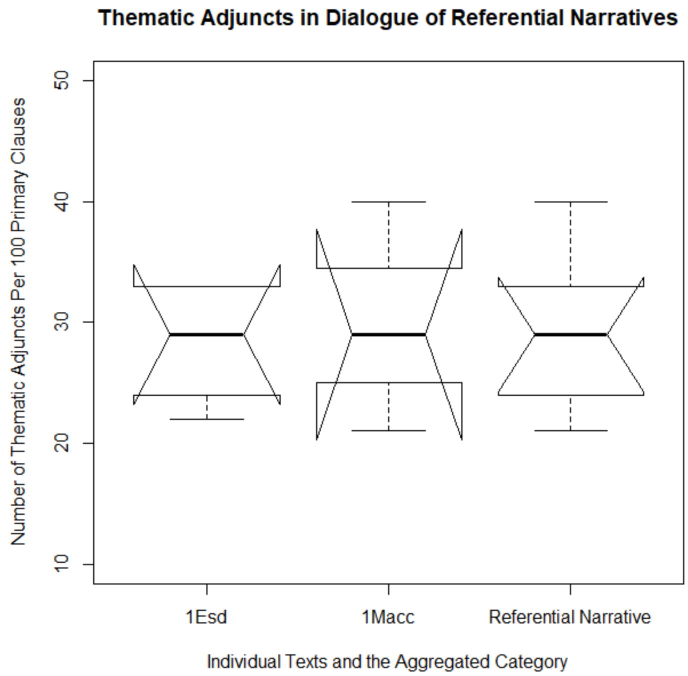
As with the framework, the data on subsystem 1 (+/- Process) does not seem to support combining the samples from 1 Esdras and 1 Maccabees. The standard deviation of the sample from 1 Maccabees is basically double that of the sample from 1 Esdras (27.98 vs. 14.01). While both samples are positively skewed, the sample from 1 Esdras is almost exactly three times that of the sample from 1 Maccabees. Of the basic descriptive statistics, only kurtosis comes close to matching; both samples have flatter peaks than the normal distribution. The confidence intervals based on the raw frequencies do not quite overlap, though they come within 1 selection/100 instances of the system of doing so. The boxplot for these data is below.



The figure above illustrates the problems with these samples. Although the confidence intervals for the sample medians overlap substantially, this is largely because the confidence interval for the median of the sample from 1 Maccabees is so wide, protruding far beyond both the lower and upper edges of the box. The low side of the box or 1 Esdras also protrudes beyond its box. The same is true for the aggregated data. Taken together, these observations indicate that, if these two samples belong to the same population, the population is a quite diffuse one. As with several other sets of data, this one could benefit from both additional data from the texts already sampled to clarify the status of these texts and samples from additional texts to clarify whether a single, dispersed population or multiple populations are represented here.

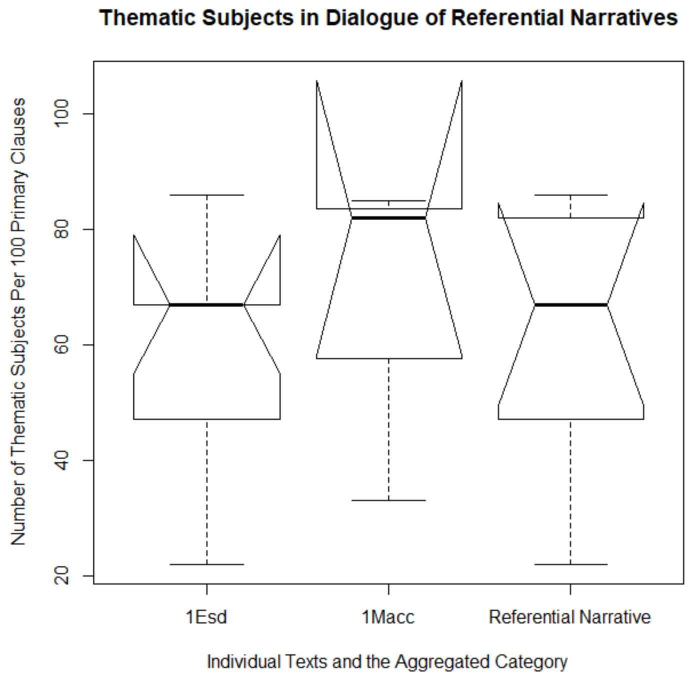
Ascertaining whether the data on subsystem 2 (+/- Circumstance) from the projected discourse portions of 1 Esdras and 1 Maccabees are compatible with one another is somewhat difficult because the indicators point in different directions. On the one hand, the basic descriptive statistics are quite different: the standard deviations are different, the kurtosis of the samples fall on differing sides of the normal distribution, and, although both samples are positively skewed, the magnitude of that skew is quite different. On the other hand, the confidence intervals based on the relative frequencies are almost identical; the wider interval subsumes the narrower one. The boxplot of these data is below.[2]

[2] In the interests of making the plot easier to read, I overrode the default settings for the scale of the y-axis. This resulted in one outlier (a normalized frequency of 63 selections/100 instances of the system for the first chapter of 1 Esdras) not appearing in the plot, but had I used the scale necessary for it to appear, some of the protruding triangles would not have been visible.



This plot presents strong evidence that the two samples belong to the same population, i.e. genre, or at least populations that have the same value for the normalized frequency of +Circumstance. Notice that the dark line marking the sample median of each plot forms a straight line, meaning that the medians are very close to each other and, as a result, the median of the aggregated data is close to the same value. This goes far beyond the minimum requirement for ascertaining whether two samples are compatible with the hypothesis that they belong to the same population. Whereas simply having the range of values not significantly different from the median of each sample overlap is sufficient to establish the compatibility of two samples, in this case the *actual* sample estimates for the median of the normalized frequency for +Circumstance are precisely equal (29 selections/100 instances of the system). The only way these two samples could turn out to be incompatible is if the sample medians are poor enough estimators of their respective population medians that they appear to be equal while that would not be true of a reliable estimate. Although the boxes in the above figure representing both samples have at least one set of protruding triangles that mark confidence intervals broader than their respective boxes and, thus, potential problems with sample representativeness, none of the confidence intervals protrude beyond the box enough to suggest that the sample estimate is sufficiently unreliable to drag the true population median far enough away from the sample median that the true median would be different to a statistically significant degree. In fact, none of the problematic confidence interval boundaries are broader than the edge of their respective box by more than 5 selections/100 instances of the system, and two of them do so by less than 1 selection/100 instances of the system. In summary, although the descriptive statistics indicate that more data should be collected, both the confidence intervals for the mean relative frequency and the median normalized frequency strongly indicate that the data already collected is sufficient to show that the two samples are consistent with the hypothesis of common genre.

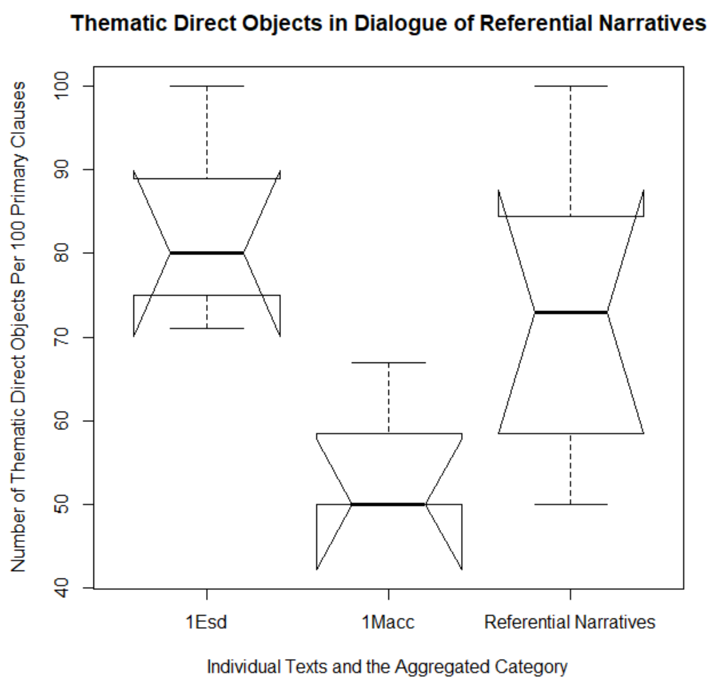
Turning to subsystem 3 (+/- Subject), the data seem to support the hypothesis of common genre. The standard deviations of the samples differ by 7 selections/100 instances of the system. Both samples are negatively skewed, though the sample from 1 Maccabees is substantially more so. Both samples have a flatter peak than the normal distribution, though the sample from 1 Esdras is substantially more so. The confidence intervals for the grammatical probability of +Subject calculated from the raw frequencies overlap by roughly 9 selections/100 instances of the system. A boxplot with confidence intervals for the median normalized frequency is below.



Once again, the plots indicate that I need to extend the samples in order to be reasonably confident that the sample median accurately represents the population median. Notably, only the upper bounds of the confidence intervals for the three medians extend beyond the boxes; the lower bounds for all three are within the box, with the boxes for 1 Esdras and the aggregated data having confidence intervals that are discernibly narrower than their respective boxes. The magnitude by which the confidence intervals exceed the size of the box varies: almost 12 selections/100 instances of the system in the case of 1 Esdras, 22 selections/100 instances of the system (into the range of impossible values, over 105 selections/100 instances of the system) in the case of 1 Maccabees, and between 2 and 3 selections/100 instances of the system in the case of the aggregated data. Clearly, additional data from both 1 Esdras and 1 Maccabees are necessary.

The data as they currently stand are more consistent with grouping 1 Esdras and 1 Maccabees than with separating them. The lower bounds for the confidence intervals of their medians are very close (roughly 55 and 58 selections/100 instances of the system, respectively). While the upper bounds for the confidence intervals have the problems described above, the observed sample median for 1 Maccabees almost falls within the upper bound of the narrower confidence interval as is, with the upper bound of values consistent with the data from 1 Maccabees being roughly 79 selections/100 instances of the system and the observed median from 1 Maccabees being 82 selections/100 instances of the system. The median of the data from 1 Maccabees is more likely to come down than go up, since the confidence interval currently includes values over 100 selections/100 instances of the system. This inference indicates that the true median for texts belonging to the same genre as 1 Maccabees, whatever that genre might happen to be, is less than the current observed value of 82 selections/100 instances of the system, implying in turn that, assuming the sample median of 1 Esdras remains relatively stable, the current margin by which the median for the data from 1 Maccabees falls outside the notch for the data from 1 Esdras will also decrease. Obviously, I cannot know ahead of time that the sample median for the data from 1 Esdras will remain relatively constant, but it seems more likely to do so (note the relative size of the notches). As a result, I conclude that these data are relatively consistent with treating 1 Esdras and 1 Maccabees as a coherent group.

As with the framework, the data on the use of subsystem 4 (+/- Affected) in projected discourse are problematic because of the rarity of thematic Complements. Nevertheless, the basic descriptive statistics are reasonably similar: the standard deviations differ by roughly 2 selections/100 instances of the system, both are positively skewed (though the sample from 1 Maccabees is more so), and both have a flatter peak than the normal distribution (though, once again, the sample from 1 Maccabees is more so). The confidence intervals for the true grammatical probability of +Affected overlap by over 25 selections/100 instances of the system, but this is only the case because the interval for the 1 Maccabees sample includes over a third of the theoretically possible values. A boxplot with the confidence intervals for the median of the normalized frequencies is below.

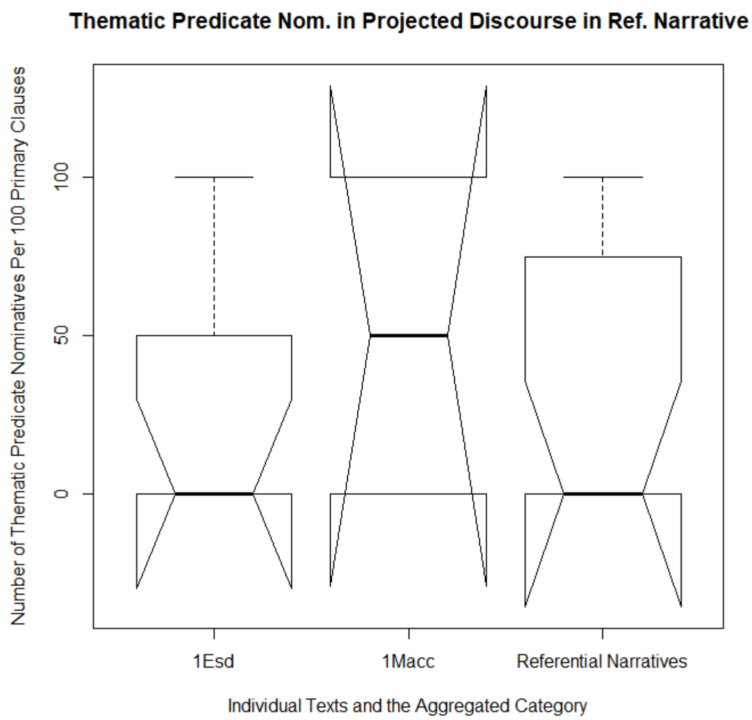


The above boxplot visualizes the frequencies with which the dialogue portions of the texts classified as referential narratives choose an accusative Complement, i.e. a direct object in traditional terms, as ideational theme relative to the frequency with which they opt for another sort of Complement (subsystem 4). This plot indicates problems with both sample length and sample depth.

All three boxes have the protruding triangles that mark problems with sample length, but the margin by which this is true varies widely. The lower bound of the confidence interval for 1 Esdras's median extends beyond the box by almost 5 selections/100 instances of the system; the upper bound, however, does so by less than 1 selection/100 instances of the system. The lower bound for the confidence interval for the data from 1 Maccabees by about 8 selections/100 instances of the system, while the upper bound falls just inside the box.

These data, as they currently stand, do not support combining 1 Esdras and 1 Maccabees. The gap between the upper bound of the confidence interval for the median of the data from 1 Maccabees and the lower bound of the confidence interval for the median of the data from 1 Esdras is roughly 12 selections/100 instances of the system. This gap indicates that, if 1 Esdras and 1 Maccabees belong to the same population, they represent opposite tails of a population distribution whose median would likely fall somewhere between the sample medians for 1 Esdras and 1 Maccabees. The way to test this is to take more samples and see if they cluster between 1 Esdras and 1 Maccabees.

The data for subsystem 5 (+/- Stative) are even more problematic with regard to sample size than the data for subsystem 4. I will include a discussion of them in the interests of completeness, however. The standard deviations differ by 13.25 selections/100 instances of the system. The skewness of the samples differs, with the sample from 1 Maccabees having no skew and the sample from 1 Esdras being positively skewed. Both samples have a much flatter peak than the normal distribution. The confidence interval for both estimates of the true grammatical probability of +Stative overlap by default because every theoretically possible value, as well as an equally large range of theoretically impossible values, lies within the confidence interval derived from the 1 Maccabees sample. A boxplot with confidence intervals for the median of normalized frequencies is below.



This figure clearly testifies to the shortcomings of this group of data as an indicator of whether the two sampled texts belong to the same population, i.e. genre category. The confidence intervals overlap, but that is because every theoretically possible value, along with many theoretically impossible ones, lies within the confidence interval for the median of the normalized frequencies in the sample from 1 Maccabees. A full half, even of the aggregated data, is at 0 selections/100 instances of the system, which cannot be an accurate representation of any population because it would mean that the population did not exist.

Summary

Five subsystems and two sets of data (framework and projected discourse) would equal ten groups of data. Unfortunately, however, subsystems 4 and 5 are too infrequent in both framework and projected discourse to address the question clearly, leaving six groups of data. Four of these groups (the other two being subsystem 1 in both the framework and projected discourse) were compatible with the hypothesis that 1 Esdras and 1 Maccabees belong to the same genre. Although, strictly speaking, being unable to prove a significant difference does not necessarily mean that the samples are similar, the degree to which several groups of data exceeded the minimum requirement of no significant difference indicates that samples are substantially similar and, thus, the statistical calculations in Chapter 7 that rely on aggregating the data from 1 Esdras and 1 Maccabees to form a picture of what I call "referential narrative" are likely to be accurate.

Thematization in Referential Narrative: Overview

The tables below report the overall frequency counts observed in the samples from referential narrative. These counts, in contrast to the ones in the following sections, are of surface realizations in the text, without reference to the underlying system. Thus, there are four columns—one for each of the clause constituent types possibly serving as ideational theme. As both these data and the data from Mark show, the narrative framework and projected discourse need separate consideration.

Table 5.3 Overview of Thematization in 1 Maccabees

	Circumstances	Process	Subject	Complement	Total
Narrative	26	316	37	15	394
Projected	17	59	33	7	116
Total	43	375	70	22	510

Table 5.4 Overview of Thematization in 1 Esdras

	Circumstances	Process	Subject	Complement	Total
Narrative	56	126	65	12	259
Projected	35	60	43	36	174
Total	91	186	108	48	433

Steps Forward

These figures represent a necessary first step towards a picture of the thematization system of referential narrative, but they are not sufficient in and of themselves. The mathematical warrant for quantitative analysis of a system rests on the paradigmatic nature of language. Only in a closed system can one meaningfully assign probabilities to individual terms. Thus, before moving on to quantitative analysis of these data, they need to be transformed from absolute frequency counts of surface realizations into counts of the systemic choices that produced them.

The first three columns of Tables 5.3 and 5.4 directly represent the output of one of the subsystems of the theme system described in chapter 3; the final column represents the output of both subsystems 4 and 5 because I only realized I needed to divide them after I began collecting the data. These outputs represent one of the choices in a subsystem, namely the choice to insert that particular clause constituent type as ideational theme. The other choice in each case is to not select that particular type of clause constituent as ideational theme but continue down the system network to select another sort of clause constituent. As a reminder, the hierarchy of subsystems is: subsystem 1 (+/- Process), subsystem 2 (+/- Circumstance), subsystem 3 (+/- Subject), subsystem 4 (+/- Affected), and subsystem 5 (+/- Stative). Thus, for example, the witness 1 Maccabees bears to the paradigmatic arrangement of subsystem 1 in narrative frameworks is 316 choices of +Process (the observed frequency) and 78 choices of -Process (the combined count of other choices). Iterating this process for both samples and all subsystems produced Tables 5.5 and 5.6; a corresponding one produced Tables 5.8 and 5.9.

Thematization in Referential Narrative: Narrative Framework

Table 5.5 Frequencies of Thematization Choices in 1 Maccabees (Framework)

Subsystem	Plus Feature	Minus Feature	Total
1 (+/- Process)	316 (0.80)	78 (0.20)	394
2 (+/- Circumstance)	25 (0.32)	53 (0.68)	78
3 (+/- Subject)	37 (0.70)	16 (0.30)	53
4 (+/- Affected)	14 (0.88)	2* (0.22)	16
5 (+/- Stative)	1* (0.5)	1* (0.5)	2

Table 5.6 Frequencies of Thematization Choices in 1 Esdras (Framework)

Subsystem	Plus Feature	Minus Feature	Total
1 (+/- Process)	126 (0.49)	133 (0.51)	259
2 (+/- Circumstance)	56 (0.42)	77 (0.58)	133
3 (+/- Subject)	65 (0.84)	12 (0.16)	77
4 (+/- Affected)	10* (0.83)	2* (0.17)	12
5 (+/- Stative)	2* (1.00)	0* (0.00)	2

Table 5.7 Normed Aggregate Frequencies (Framework)

Subsystem	Plus Feature (per 100 Clauses)	Minus Feature (per 100 Clauses)
1 (+/- Process)	67 (0.67)	33 (0.33)
2 (+/- Circumstance)	38 (0.38)	62 (0.62)
3 (+/- Subject)	72 (0.72)	28 (0.28)
4 (+/- Affected)	88 (0.88)	12 (0.12)
5 (+/- Stative)	75 (0.75)	25 (0.25)

An Overall Picture of Thematization in the Framework of Referential Narrative

The last subsection of the previous section laid out the rationale that produced Tables 5.5 and 5.6; now I turn to explaining Table 5.7. Producing an overall picture of ideational theme choice in referential narratives involves combining the samples from 1 Maccabees and 1 Esdras. These samples are of unequal lengths, meaning that comparing them in absolute terms would be meaningless. Thus, I have computed the normalized frequency per 100 clauses. Computing the normalized frequency consists of determining what percentage of the overall sample came from each individual sample and weighting the sample mean accordingly. Since the scale I chose for normalization was 100 clauses, the proportions are simply the frequencies divided by 100.

Thematization in Referential Narrative: Projected Discourse

Table 5.8 Frequencies of Thematization Choices in 1 Maccabees (Discourse)

Subsystem	Plus Feature	Minus Feature	Total
1 (+/- Process)	59 (0.51)	57 (0.49)	116
2 (+/- Circumstance)	17 (0.30)	40 (0.70)	57
3 (+/- Subject)	33 (0.71)	7* (0.29)	41
4 (+/- Affected)	4* (0.57)	3* (0.43)	7
5 (+/- Stative)	3* (1.00)	0* (0.00)	3

Table 5.9 Frequencies of Thematization Choices in 1 Esdras (Discourse)

Subsystem	Plus Feature	Minus Feature	Total
1 (+/- Process)	60 (0.34)	114 (0.66)	174
2 (+/- Circumstance)	35 (0.31)	79 (0.69)	114
3 (+/- Subject)	43 (0.54)	36 (0.46)	79
4 (+/- Affected)	28 (0.78)	8 (0.22)	36
5 (+/- Stative)	3 (0.38)	5 (0.62)	8

Table 5.10 Normed Aggregate Frequencies (Discourse)

Subsystem	Plus Feature (Per 100 Clauses)	Minus Feature (Per 100 Clauses)
1 (+/- Process)	41	59
2 (+/- Circumstance)	30	70
3 (+/- Subject)	60	40
4 (+/- Affected)	78	22
5 (+/- Stative)	54	46

An Overall Picture of Thematization in Projected Discourse of Referential Narratives

The three preceding tables follow the same format as the analogous ones for the narrative framework, presented in the previous section. The non-parenthetical numbers in Tables 5.8 and 5.9 are the raw counts from projected discourse for the various options in the clause thematization system. In each case, the number in the "Plus Feature" column corresponds to the observed frequency of the corresponding type of clause constituent: for instance, there are 59 thematic verbs within the projected discourse portions of 1 Maccabees and, thus, 53 instances of +Process. The non-parenthetical number in the "Minus Feature" column of both tables corresponds to the sum of the "Plus Feature" results for the rows below. In each case, the parenthetical numbers are the observed counts expressed as a percentage of the total number of choices. Table 5.10 weights the corresponding cells of Tables 5.8 and 5.9 according to the relative lengths of the samples from 1 Maccabees and 1 Esdras.[3]

The Shape of Referential Narrative's Thematization System

These data raise analytical questions regarding thematization choice in referential narrative. The first subsection describes patterns observed in the framework portions; the second subsection describes patterns observed in the projected discourse portions. Both subsections begin by attempting to discern a basic pattern for the environment in question and then proceed by suggesting a possible functional explanation for deviations from this pattern.[4]

---

[3] See "Lessons Learned" in Chapter 8 for a better way to handle this aspect of the procedure in the future.

[4] Halliday, "Linguistic Function," 88–125.

Another way of expressing this is finding the unmarked syntax and exploring the additional meaning carried by marked syntax. Porter suggests four hallmarks of unmarked syntax, with the hallmarks of marked syntax simply being the converse of these: (1) higher frequency counts, (2) more irregular structure, (3) fewer expressed constituents, and (4) "the minimum essential meaning."[5] The first three of these are easy to quantify; the fourth is more qualitative, essentially being a case where the proof is in the pudding: if I can consistently explain significant deviations from a pattern, a strong case exists that the pattern in question is the unmarked, or default, choice.

A significant deviation is defined as a data point that falls more than two standard deviations from the mean for the group of which it is a part. In the case of my project, the data points in question are the normalized frequencies from each chapter, which form the basis of the means and standard deviations reported earlier. The discussions below report all cases where the normalized frequency for a choice met this standard and seek functional explanations within the content of the chapter for why the author chose different thematization patterns from those manifested elsewhere within that text.

Few divergences at the chapter level are large enough to constitute a significant deviation by this standard. This can be considered a positive state of affairs in some ways because, if the standard deviation is low, it implies that the mean normalized frequency for the text as a whole reliably estimates the mean of the population as a whole, which in turn implies that the t-tests in Chapter 7 based on this mean normalized frequency will yield reliable results. On the other hand, a large standard deviation, which indicates uncertainty regarding the accuracy of the mean normalized frequency as an estimate for

---

[5] Porter, "Word Order and Clause Structure," 190.

the mean of the population as a whole, may prevent the discovery of significant deviations simply because twice the standard deviation on either side of the mean encompasses the majority of possible values and, thus, the values that actually showed up in the sample. As usual, the remedy for this situation would be to collect more data: sample size is the denominator of the formula for standard deviation, so more data points will naturally lower the standard deviation. I covered the accuracy of the sample estimates earlier in the chapter; I bring it up again simply to point out that more accurate sample estimates more clearly display significant deviations. Nevertheless, I press on with suggesting unmarked and marked thematization patterns suggested by the data as they currently stand.

Trends in Thematization Choice in the Frameworks of Referential Narrative

Clearly, the unmarked pattern in the frameworks of referential narrative is to place the verb in thematic position (i.e. choose +Process in subsystem 1). This pattern accounts for just over 80 percent of the clauses in the framework portion of the sample from 1 Maccabees; although the proportion is lower in the case of 1 Esdras, thematic verbs, realizing +Process, still outnumber their nearest competitor by roughly two to one. The divergence in proportion between 1 Maccabees and 1 Esdras bears witness to the second of Porter's criteria for unmarked syntax, namely irregularity of structure. In terms of the third criterion, bulk, primary clauses consisting only of a coordinating conjunction and verb are much more common than clauses consisting of only one of the other three types of clause constituent. Evaluating whether this pattern meets the fourth criterion, "the minimum essential meaning," is difficult because none of the normalized frequencies from the individual chapters of either text met the standard of a significant deviation, meaning there are no chapters to mine for functional explanations for the use of -Process that could suggest the additional implicatures associated with marked syntax. The choice to place a verb in thematic position fulfills three of the four criteria for unmarked syntax; the jury is still out on the fourth.

Turning to subsystem 2 (+/- Circumstance), the default, unmarked choice is -Circumstance and, thus, +Circumstance is a marked choice. The more frequent choice in both texts is -Circumstance. The estimates for the grammatical probability of +Circumstance derived from the two samples differ by only 10 percent, less than half the average deviation between estimates, indicating greater regularity in the use of +Circumstance. No clause in the samples consists of only adjuncts, whereas the other constituent types can occur by themselves, indicating that +Circumstance is associated with bulkier clauses. Lastly, though the samples provided no significant deviations on which to test the theory, the optional nature of Adjuncts as a whole suggests that their inclusion, particularly in thematic position, is significant. Thus, three of the four criteria for differentiating unmarked and marked syntax clearly point to +Circumstance as a marked choice and -Circumstance as the default choice; the same would seem to be true for the fourth criterion on theoretical grounds, though the data do not permit quantitative verification.

The choice to place a Subject in thematic position seems to be a marked choice relative to the choice to place a verb in that position but not as strongly marked as the choice to place an Adjunct or Complement there. The observed frequencies point in this direction: thematic Subjects are more common than thematic Adjuncts or Complements but less frequent than thematic verbs. The difference between the sample estimates for the grammatical probability of +Subject differ only slightly more than those for +Circumstance, suggesting a regularity of structure. Expressed subjects tend to be associated with more complicated clauses. None of the chapter-by-chapter normalized frequencies differed from the mean by more than two standard deviations, so there is no material with which to develop a functional explanation, though previous research has suggested that thematic Subjects signal "topic shifts as new characters and new subjects of discussion."[6] Nevertheless, three of the four criteria support the markedness of +Subject.

Placing accusative Complements in thematic position seems to be a quite marked choice. The frameworks of the two samples, a total of 653 primary clauses, contain a combined total of 24 instances of +Affected. The difference between the sample estimates for the grammatical probability of +Affected is the smallest of any subsystem, pointing to a very regular structure. Accusative complements rarely occur as the sole constituent of a clause, suggesting that thematic Complements are associated with longer, more complex clauses.

Although none of the chapter-by-chapter frequencies meet the standard of a significant deviation, they are not spread evenly across the chapters: several chapters have none, but only one of the chapters that has them (1 Esdras 9) does not have multiple. This suggests that thematic accusative Complements tend to occur in clumps. First Maccabees 2 illustrates this. The chapter contains four thematic direct objects, two of which occur next to each other in verse twenty-five. In terms of functional explanations, this clump seems to be emphasizing the significance of the entities Matthias was

---

[6] Porter, "Word Order and Clause Structure," 203.

destroying, killing a royal official and getting rid of the altar appointed for idolatrous sacrifices, a conclusion reinforced by the reference to Phineas in the next verse. One of the remaining instances fronts the Complement to emphasize that the Jews were not willing to break the Sabbath at all, even to "hurl a stone" (2:36); the last instance is a rare instance of ellipsis in the narrative framework, though even here the Complement is not alone, being joined by an Adjunct (2:44).

Both options in subsystem 5 (+/- Stative) seem marked relative to the choice to place a verb in thematic position, but the presence of a thematic indirect object, realizing the choice -Stative, seems especially significant. The framework portions of these two samples contain a grand total of three selections from this subsystem, two of which are +Stative, and one of which is -Stative. With so few instances, it is impossible to make any determinations about regularity of structure or bulk of clauses. One of the thematic predicate nominatives, realizing the choice +Stative, seems to emphasize the success of the rebuilding of the Temple (1 Esdras 7:3). The one thematic indirect object (1 Maccabees 2:2) introduces the family of Matthias, who are the focus of the book.

Trends in Thematization Choice in Discourse Within Referential Narrative

The data in Tables 5.8 through 5.10 indicate that +Process is no longer the default choice when looking at projected discourse. The culprit seems to be a shift toward the categories that nouns typically realize, i.e. Subject and the various forms of Complement, for the relative frequency of thematic adjuncts remains relatively stable across the two environments, ranging from 30 percent of the choices for subsystem 2 (+/- Circumstance) in the projected discourse portions of 1 Maccabees to 42 percent in the framework of 1 Esdras. Since subsystem 2 is relatively stable, deviations in the relative frequencies of the terms in subsystems 3 through 5 must account for the increased relative frequency of -Process in projected discourse. In fact, the bulk of the difference appears to come from subsystem 4 (+/- Affected). I suggest that an increase in ellipsis within projected discourse may account for this: whereas, as noted above, ellipsis is rare within the framework, it is not uncommon within projected discourse, particularly in responding to questions.

Having laid the groundwork of the basic pattern within the projected discourse portions of referential narrative, the way is now paved for considering the portions of each sample that stand out against the rest of that sample. Standing out from the rest of the sample is defined as being more than two standard deviations away from the mean. Two groups of data qualified, namely the use of subsystem 2 (+/- Circumstance) in 1 Esdras 1 and the use of subsystem 3 (+/- Subject) in 1 Esdras 9.

The normalized frequency for +Circumstance in 1 Esdras 1 is 63 selections/100 instances of the system, which is roughly one-half of one selection/100 instances of the system less than two standard deviations away from the mean. Thematic adjuncts occur five times in this chapter; all of them in significant contexts. The first instance is Josiah noting that it was time for the Levites to begin worship (1:4). The second is the following: καὶ στάντες ἐν τῷ ἱερῷ κατὰ τὴν μεριδαρχίαν τὴν πατρικὴν ὑμῶν τῶν Λευιτῶν τῶν ἔμπροσθεν τῶν ἀδελφῶν ὑμῶν υἱῶν Ισραηλ ἐν τάξει θύσατε τὸ πασχα (1 Esdras 1:5–6). The aorist participle στάντες is functioning as the Predicator of an embedded clause that is a thematic Adjunct, describing where the priests are to be when they offer the Passover sacrifice. Note that the embedded Predicator itself is further qualified by Adjuncts: Josiah is being very specific about how he expects the sacrifice to be conducted in accordance with the Law of Moses. The other three instances occur in rapid sequence in verse twenty-five, making three arguments for Josiah staying out of Pharaoh's conflict: (1) God sent Pharaoh to Josiah, (2) Pharaoh's war is at the Euphrates (and, hence, not against Judah), and (3) God is currently backing Pharaoh (implying that Josiah should back off, lest he find himself fighting against God). Thus, each instance of a thematic adjunct in 1 Esdras 1 occurs at an argumentatively significant point.

The estimated grammatical probability for +Subject in 1 Esdras is 54 selections/100 instances of the system. The normalized frequency for 1 Esdras 9 is 86 selections/100 instances of the system, which is more than two standard deviations above the mean, indicating a significant shift towards Subjects and away from the various sorts of Complements. The one thematic Complement in the chapter is interesting because it is a direct repetition of something one of the leaders said previously: when Ezra finishes reading the Law, Attarates says that "the day is holy to the Lord" (S-C order); when the Levites repeat this, they use C-S order. Making an iron-clad case from an isolated occurrence is difficult, but I think it is likely that the Levites are stressing the inappropriateness of weeping, the initial response to Ezra reading the law.

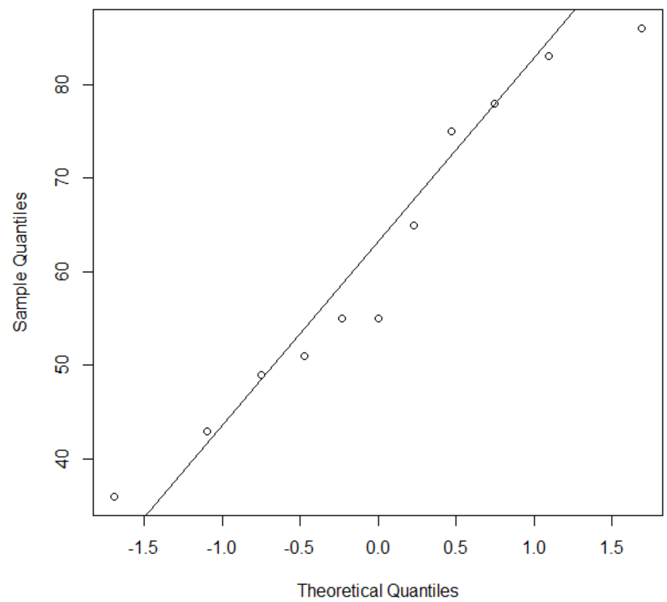
Are the Referential Narratives Normal?

The next general consideration in dealing with referential narratives is what statistical analyses are appropriate for analyzing the data derived from them. As with the corresponding discussion of Mark, this boils down to asking how closely the data conform to the normal distribution, since that is the intractable assumption of a t-test. The first subsection discusses the normality of the data from the frameworks of referential narrative. The second subsection discusses the normality of the data from the dialogue portions of referential narrative. The third subsection summarizes the import of these results for the statistical tests found in chapter 7.

As a reminder, testing a group of data for normality consists of several steps. The first step is to generate a quantile-quantile (Q-Q) plot. Quantiles divide a dataset into five parts, each representing 20 percent of the data. The reference line on my Q-Q plots shows where the quantiles would be, if the data perfectly matched the normal distribution. Thus, if the dots representing the quantiles of the sample cluster around the reference line, then the sample's distribution is close to the normal distribution. The second step is to calculate the Pearson correlation coefficient (designated r) as a measure of the relationship between the data sample and the normal distribution. The third step is to look at a histogram of the data and compare it to the bell-shaped curve of the normal distribution. This step is important because sets of data with balanced peaks on extreme ends of the distribution and almost no data in the middle of the plot, a distribution referred to as the bimodal distribution, will return a high value for r because the peaks on the extremes will cancel each other out. Despite the supposed correlation, a t-test on bimodal data would be meaningless because the t-test assumes that the data cluster around the central value. The value of this central value can be estimated in one of three ways: (1) adding up all the values and then dividing by the number of values, i.e. the mean, (2) listing the values from least to greatest and determining what the center value is, i.e. the median, or (3) determining the most common value, i.e. the mode. The last step of testing a sample for normality is to determine all three of these values and compare them. The values of all three would be equal if the data are perfectly normal. Thus, if one or more of these values, collectively called the measures of central tendency, differ greatly from the others the data do not match the normal distribution.

Framework Normality

Figure 1 below displays the Q-Q plot for the data collected on subsystem 1 from the



frameworks of referential narratives. The data do not cluster around the reference line as closely as the corresponding plot for Mark, but only a few of the points are substantially off the line. The value computed for r is 0.98. Figure 2 below is the histogram for the same data. Note that the histogram falls off quicker on the right than it does on the left but then rises again. Nevertheless, there is still one peak, and it is relatively close to the center. The median and mode are both 55, while the mean is 61. The skewness value is 0.12, i.e. the data are almost symmetrical. The distribution has a flatter top than the normal distribution; the normal distribution's kurtosis is 1.32 higher than that of this data set. While the data are not ideal, they are clearly close enough for t-tests to be meaningful.

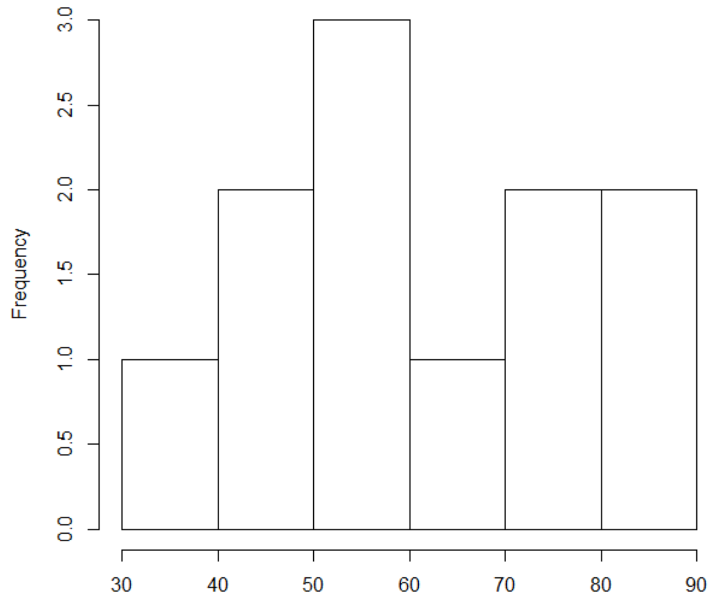
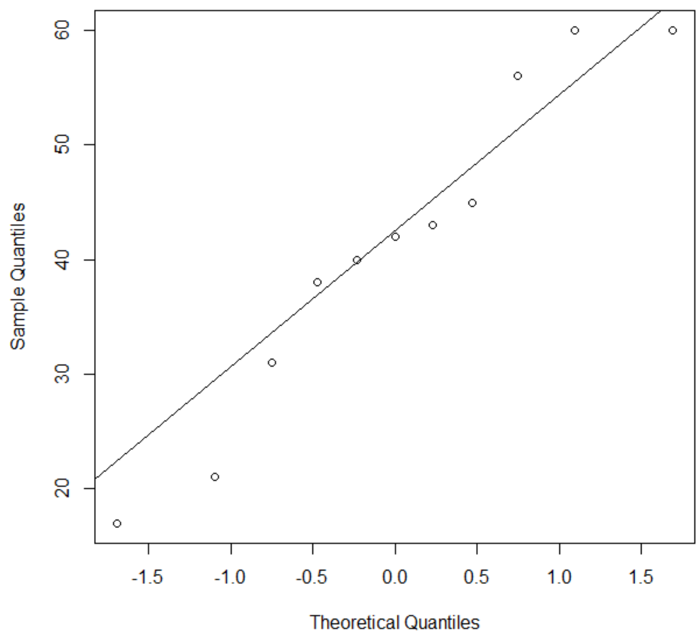
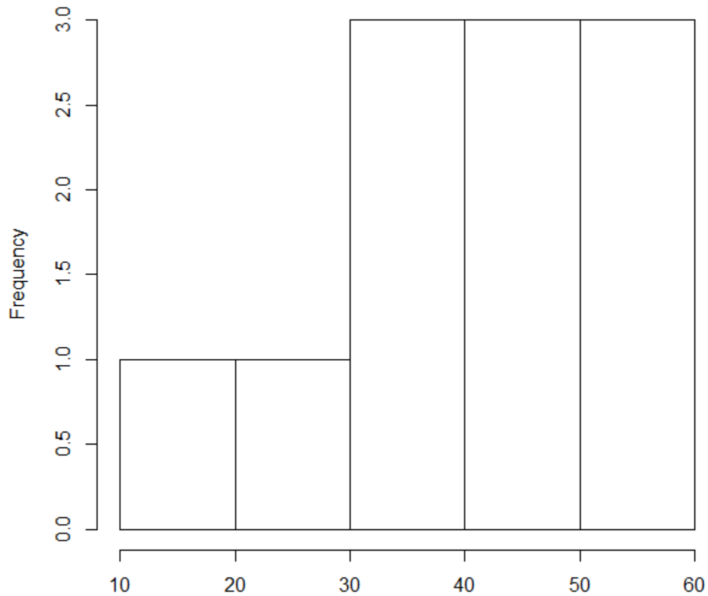


Figure 3 below shows the Q-Q plot for the data on subsystem 2 (+/- circumstance) taken from the frameworks of referential narrative samples. This plot shows discernible clustering around the reference line in the center of the plot, but several divergent values



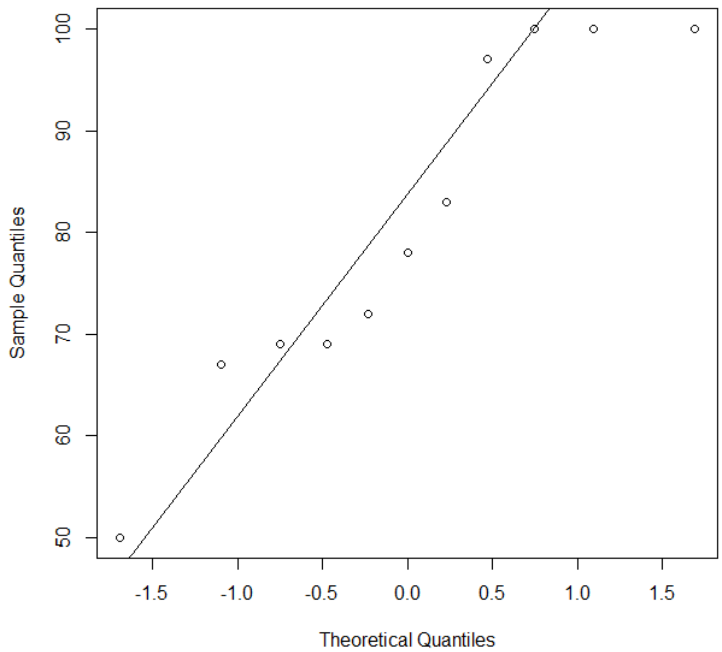
exist on the edges.

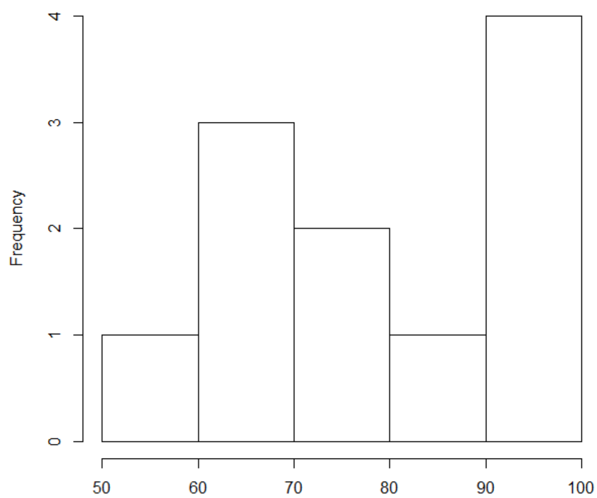
The correlation coefficient is slightly lower than for subsystem 1 but still quite reasonable (r=0.97). Unfortunately, the histogram (Figure 4) does not taper at all on the right side. This shows up in the measures of central tendency: the mean is 41, the median is 42, and the mode is 60. The skewness of the data is -0.24; the kurtosis is less than the



normal distribution's kurtosis by 0.87. Overall, these data do not support the use of the t-test.

Figure 5 below shows the Q-Q plot for the data on subsystem 3 (+/-subject) taken from the frameworks of referential narrative. Note that only one of the points falls directly on the reference line and the divergences are almost all to the right. The value computed for the correlation coefficient is lower than the ones for subsystems 1 and 2 (r=0.95). The histogram (Figure 6) shows why most of the points in Figure 5 fall to the





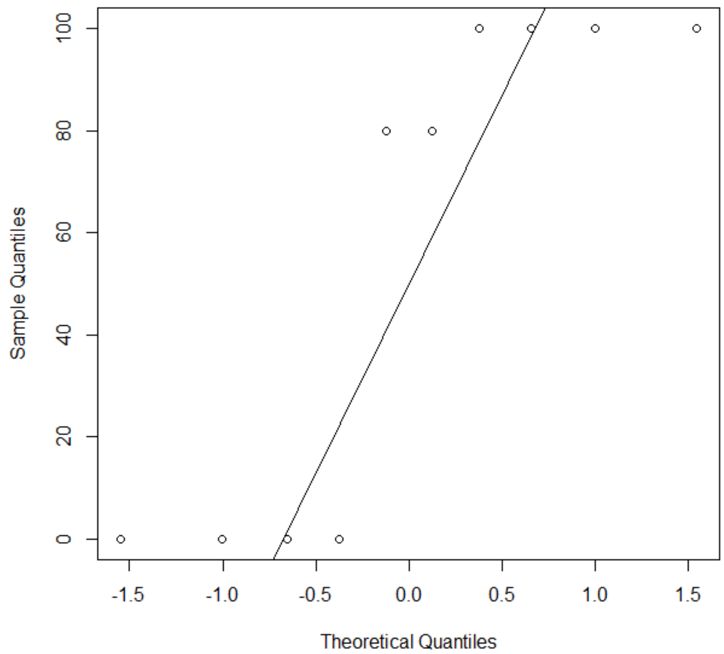
right of the reference line: the peak of the histogram is at the extreme right.

Given that this peak is the mode, pruning it as an outlier is clearly out of the question, but it is worth noting that the rest of the histogram is bell-shaped.[7] The

[7] See "Lessons Learned" in Chapter 8 for possible ways of cleaning up data like this.

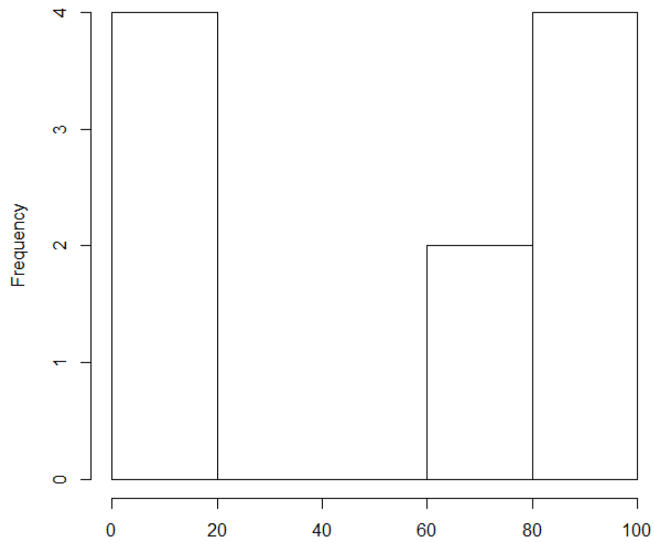
measures of central tendency reflect this: the median is 78, the mean is 80, and the mode is 100. The skewness is -0.16. The kurtosis relative to the normal distribution is -1.08. These data clearly do not support the use of a t-test.

Figure 7 shows the Q-Q plot for the data on subsystem 4 from the frameworks of

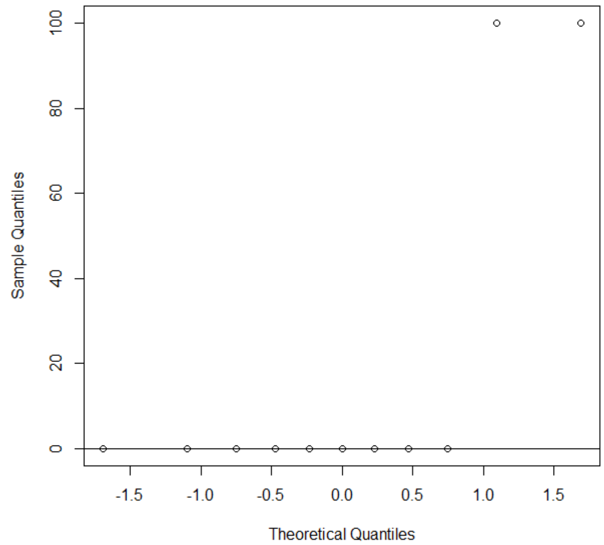


referential narrative. Note that only two of the data points fall in the middle of the diagram; this comes about because these data have very few thematic Complements, resulting in the majority of the dots being at either 0 selections/100 instances of the system, reflecting 0 instances of +affected in that chapter, or at 100 selections/100 instances of the system, reflecting that all the choices—usually only one—were

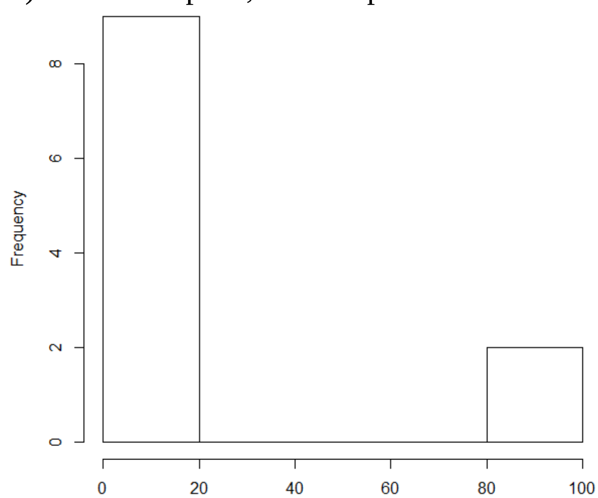
+affected. The histogram (Figure 8) reinforces this picture.



The correlation coefficient is abysmal compared to the previous one (r=0.85). The measures of central tendency diverge widely: the mean is 56, the median is 80, and 0 and 100 are tied for the mode, making this the epitome of a bimodal distribution. The skewness is -0.34; the kurtosis relative to the normal distribution is -1.80. Clearly, these data do not support the use of a t-test. For the sake of completeness, Figure 9 shows the Q-Q plot for the data on subsystem 5 from the frameworks of referential narrative.[8] Notice that the two instances where there were choices of +stative are not enough to even



cause the reference line to budge from the rest of the data. The value of the correlation coefficient is even lower than the one for subsystem 4 (r=0.70). The histogram (Figure 10) shows one peak, but that peak is at an extreme, rather than the middle. The only other



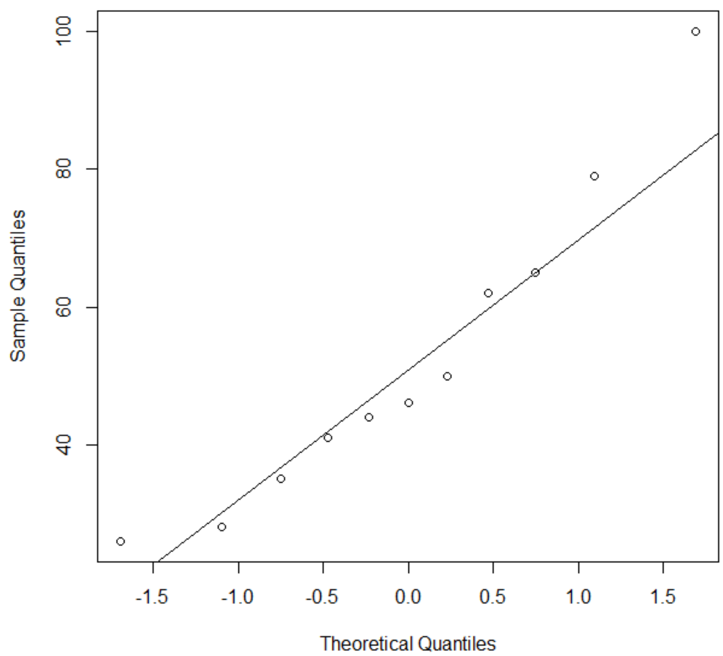
data is precisely at the other extreme.

[8] See "Lessons Learned" in chapter 8 for some ideas on procedural improvements to keep this from happening.

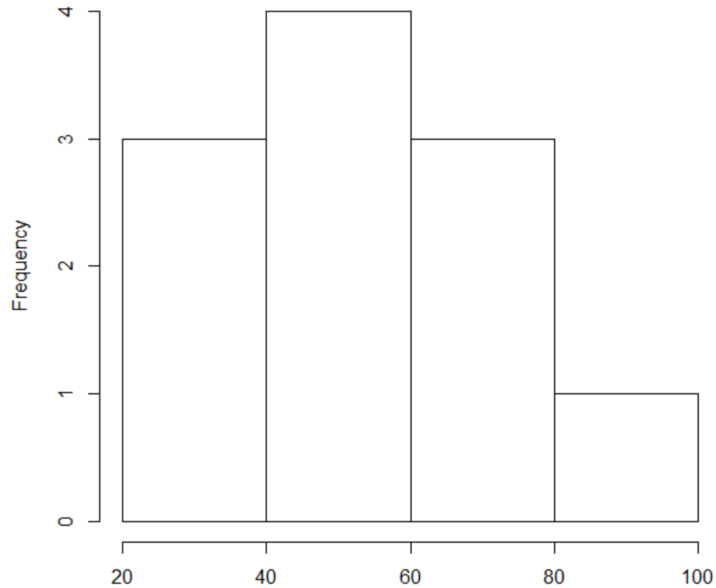
This is the opposite of a bell curve. The measures of central tendency vary significantly: the mean is 18, the median is 0, and the mode is 0. The skewness is 1.65, representing the predominance of 0 selections/100 instances of the system. This subsystem is the only one among the framework of referential narratives to be more peaked than the normal distribution, having an excess kurtosis of 0.72.

Discourse Normality

The Q-Q plot for the data on subsystem 1 in the projected discourse of referential narratives (Figure 11) shows some discernible clustering around the reference line,



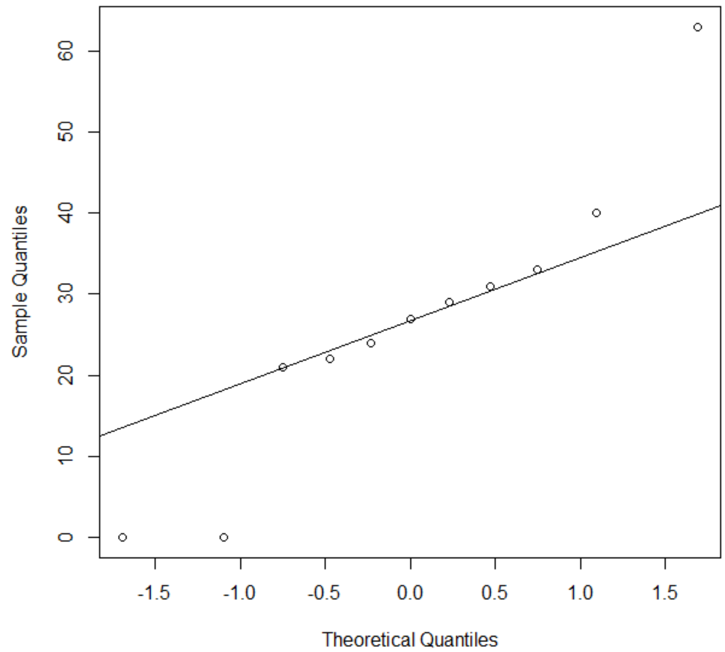
although only one data point lies precisely on it. The correlation coefficient is quite reasonable (r=0.96). The histogram (Figure 12) has one peak located near the center of



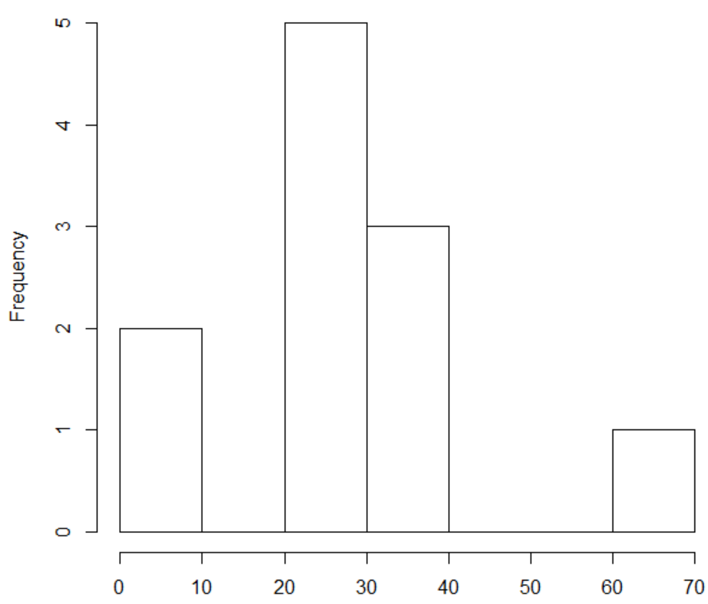
the data set.

The mean of these data is 52 selections/100 instances of the system, and the median is 46 selections/100 instances of the system. These data have no mode because each value occurred exactly once. The skewness of the data is 0.50; the kurtosis relative to the normal distribution is –1.42. While the Q-Q plot indicates room for improvement in the normality of the data, the correlation coefficient, histogram, and the measures of central tendency all support using the t-test.

The Q-Q plot for the data on subsystem 2 in the projected discourse of referential narrative (Figure 13) shows quite close clustering around the reference line: five of the data points touch the reference line—with another three nearby—while only the final

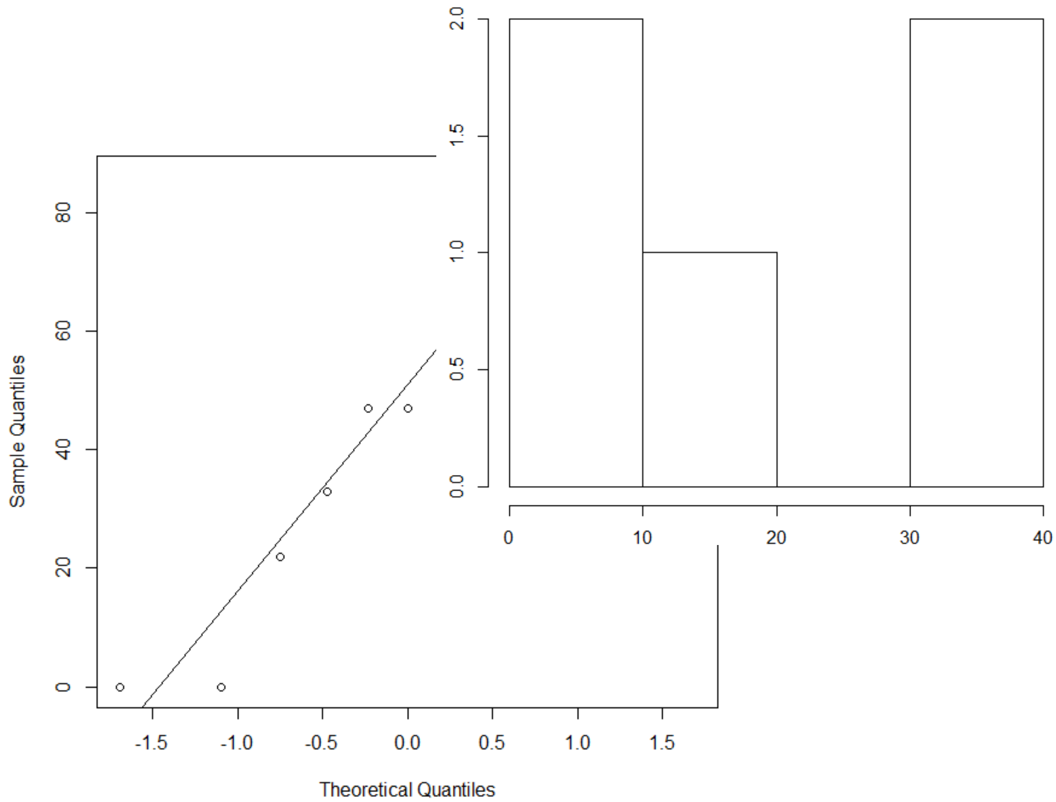


three diverge substantially from it. The correlation coefficient is well within the bounds of this study (r=0.95). The histogram of these data (Figure 14) has a single peak located near the middle of the data. The mean is 26 selections/100 instances of the system. The



median is 27 selections/100 instances of the system. These data have no mode. The data are almost completely symmetrical (skewness is -0.05). The graph of these data would be broader than the normal distribution, since the kurtosis is 1.04 less than that of the normal distribution. All these factors, except for the last one, indicate that the t-test is admissible for this group of data, and the kurtosis is just barely out of the normal range.

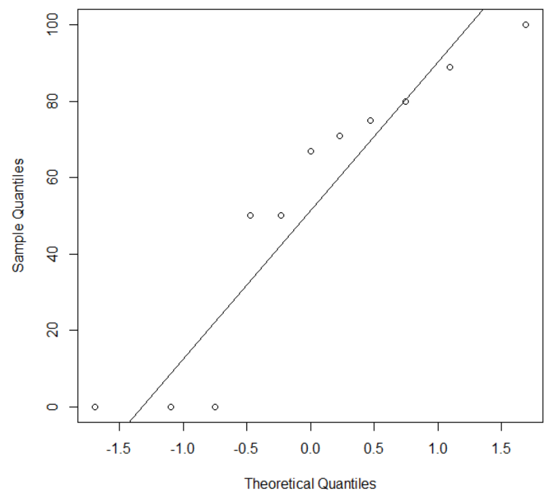
The points in the Q-Q plot for the data on subsystem 3 in the projected discourse



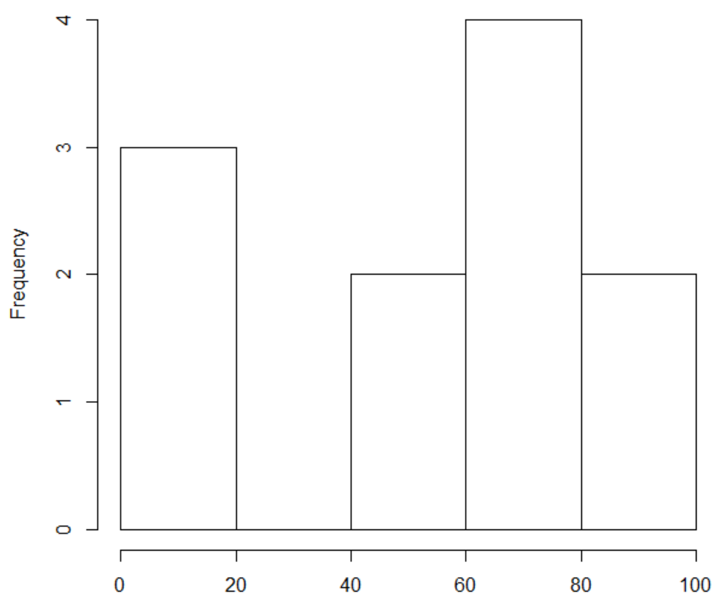
of referential narrative (Figure 15) somewhat cluster around the reference line, although—in contrast to the previous one—the line passes directly through only one point. The correlation coefficient for the relation between these data and the normal distribution is well within the limits set for this study (r=0.96). Both of these factors would seem to indicate some coherence with the normal distribution, but the histogram (Figure 16) tells a different story.

This is almost a classically bimodal histogram. The mean is 19 selections/100 instances of the system, the median is 19 selection/100 instances of the system and there is no mode. The skewness is -0.03; the kurtosis relative to the normal distribution is -

1.71. These figures indicate that the data is symmetrical like the normal distribution but markedly broader.



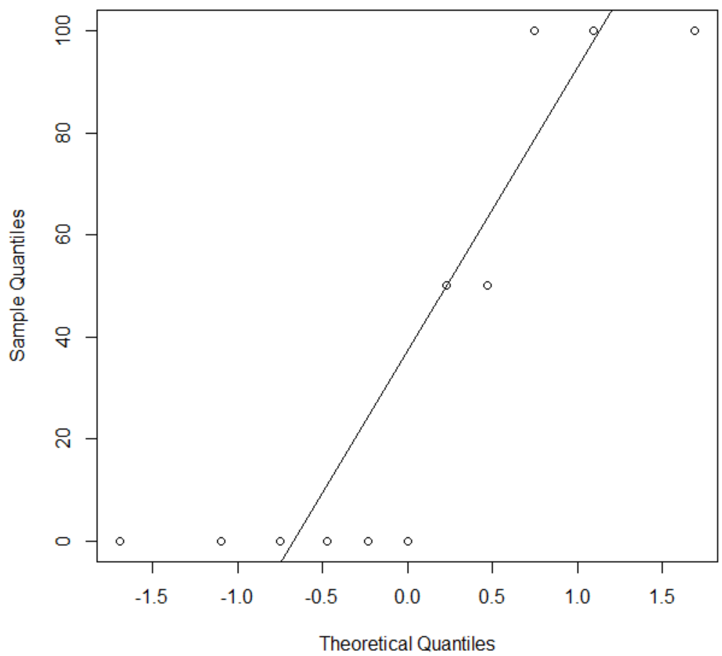
Several of the points in the Q-Q plot of the data on subsystem 4 taken from the projected discourse portions of referential narratives (Figure 17) diverge significantly from the reference line. The relationship between these data and the normal distribution is not particularly strong (r=0.93). The histogram of these data (Figure 18) is clearly bimodal; to boot, one of the modes is at the left extreme. The measures of central tendency are: 53 selections/100 instances of the system (mean), 67 selections/100



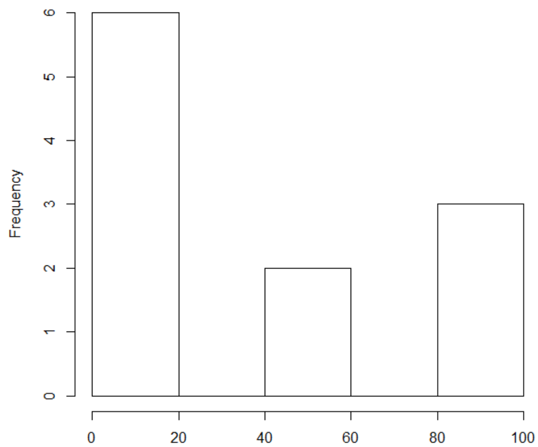
instances of the system (median), and 0 selections/100 instances of the system (mode).

All of these indicators point to not using t-tests on this data, even though the skewness (0.96) and the kurtosis relative to the normal distribution (-0.25) are both quite reasonable.

As the Q-Q plot for the data on subsystem 5 in projected discourse of referential narratives shows (Figure 19), only two data points fell anywhere other than one of the extremes, resulting in almost no clustering around the reference line. The coefficient of



correlation between this data sample and the normal distribution (r=0.87) and the histogram (Figure 20) tell the same story. The mean of this data is 36, the median is 0, and the mode is 0. The skewness is 0.56; the kurtosis relative to the normal distribution is -1.43. Although the median and mode are the same, these data do not conform to the normal distribution at all, and so t-tests on this sample would be worthless.



Summary

The tests reported above supported the use of t-tests for several groups of data taken from referential narrative. Data taken from the frameworks of referential narrative supported the use of t-tests for subsystem 1 only. The data taken from projected discourse supported the use of t-tests for subsystem 1 and 2.

The data from Mark allowed for all these t-tests, but they also allowed for t-tests on subsystem 2 in the framework and subsystem 3 in projected discourse. This disparity means my study would potentially benefit from trying to improve the normality of the data taken from referential narratives.

This subsection established that the framework data for subsystem 1 and subsystem 2 allow for reliable t-tests as they stand. Additionally, after pruning clear outliers, the data from subsystem 3 do as well. Only subsystems 1 and 2 seem normally distributed among the discourse portion of the referential sub-corpus, although potentially subsystem 4 could work as well. I will return to these results in section 7.1, where I report the actual comparisons, but first I turn to describing an overview of the data from referential narrative as a whole.

CHAPTER 6: MAPPING NONREFERENTIAL NARRATIVE

Whereas Chapter 4 discussed Mark and Chapter 5 discussed the portion of my comparative corpus devoted to referential narratives, this chapter discusses the other component of the comparative corpus, namely nonreferential narratives. The clause thematization patterns in non-referential narrative appear below.

Quantifying the Grammatical Systems of Non-Referential Narratives

Having discussed some of the problematic decisions in the annotation process, the next task is to describe the shape of the distribution of choices in the clause thematization system for the three texts I chose as representatives of non-referential narrative, treating the narrative framework and projected discourse separately since Chapter 4 showed significantly different tendencies for the two groups in the case of Mark. Thus, the first subsection below covers the framework portions of the texts I selected as representatives of non-referential narrative; the second subsection covers the projected discourse portions of these texts. These discussions report the results of applying the procedure described in Chapter 3 under the heading "Quantifying Grammatical Systems" to the samples collected from each text on a subsystem-by-subsystem and sample-by-sample format. These results are important because they are a means of assessing whether the texts, which I chose to sample on the basis of Sterling's qualitative categories, in fact form a quantitatively unified group. They also serve as part of the quantitative background against which one can position an argument for a particular pattern being significant or marked (see "The Shape of Thematization in Non-Referential Narrative" below). Lastly, these results constrain what statistical tests are appropriate for each group of data (see "Are These Data Normal?" below).

Quantifying the Frameworks of Non-Referential Narrative

Thematic verbs, realizing the choice +Process, are by far the most common choice within the framework portions of the sample from Judith. The sample estimate for the grammatical probability of +Process is 76 selections/100 instances of the system. The typical spread around this value is 8.44 selections/100 instances of the system. This sample is heavily skewed in a negative direction (-0.81). This sample has a peak somewhat flatter than that of the normal distribution (kurtosis of 2.10).

Thematic verbs are also a large component of the thematization choices within Tobit. The sample estimate for the grammatical probability of +Process is 73 selections/100 instances of the system. The typical spread around this value is 18.47 selections/100 instances of the system. This sample is somewhat negatively skewed (-0.46). This sample has a somewhat flatter peak than the normal distribution (kurtosis of 2.44).

Thematic verbs are particularly common in the sample from *Joseph and Aseneth*. The sample estimate for the grammatical probability of +Process is 89 selections/100 instances of the system. The typical spread around this value is 6.72 selections/100 instances of the system. This sample is somewhat negatively skewed (-0.41). This sample has a steeper peak than the normal distribution (kurtosis of 3.48).

Thematic adjuncts, realizing the choice +Circumstance, are the third most common category in the framework portions of Judith. The sample estimate for the grammatical probability of +Circumstance is 37 selections/100 instances of the system.

The typical spread around that value is 29.88 selections/100 instances of the system. This sample is skewed positively (0.67). This sample has a flatter peak than the normal distribution (kurtosis of 2.35).

Thematic adjuncts are also the third most common category in the framework portions of Tobit. The sample estimate for the grammatical probability of +Circumstance is 45 selections/100 instances of the system. The typical spread around this value is 36 selections/100 instances of the system. This sample is somewhat positively skewed (0.27). This sample has a much flatter peak than the normal distribution (kurtosis of 1.78).

Thematic adjuncts are also the third most common category in the framework portions of *Joseph and Aseneth*. The sample estimate for the grammatical probability of +Circumstance is 41 selections/100 instances of the system. The typical spread around this value is 22.89 selections/100 instances of the system. This sample is negatively skewed (-0.12). This sample has a flatter peak than the normal distribution (kurtosis of 2.15).

Thematic subjects are the second most common category in the framework portions of Judith. The sample estimate for the grammatical probability of +Subject is 65 selections/100 instances of the system. The typical spread around this value is 24.02 selections/100 instances of the system. This sample exhibits a slight negative skew (-0.09). This sample has a flatter peak than the normal distribution (kurtosis of 2.08).

Thematic subjects are also the second most common category in the framework portions of Tobit. The sample estimate for the grammatical probability of +Subject is 85 selections/100 instances of the system. The typical spread around this value according to the normalized frequencies is 48.70 selections/100 instances of the system. This sample is significantly skewed in a negative direction (-0.45). This sample has a much flatter peak than the normal distribution (kurtosis of 1.29).

Thematic subjects are also the second most common category in the framework portions of *Joseph and Aseneth*. The sample estimate for the grammatical probability of +Subject is 85 selections/100 instances of the system. The typical spread around this value is 36.44 selections/100 instances of the system. This sample is radically skewed in the negative direction (-1.41). This sample has a sharper peak than the normal distribution (kurtosis of 3.61).

Thematic Complements of all sorts are relatively infrequent in the framework of Judith, but direct objects, realizing the choice +Affected, are more common than either of the categories involving subsystem 5 (+/- Stative). The sample estimate for the grammatical probability of +Affected is 92 selections/100 instances of the system. The typical spread around this value is 46.29 selections/100 instances of the system. This sample is quite negatively skewed (-1.15). This sample has a flatter peak than the normal distribution (kurtosis of 2.33).

Thematic Complements are also infrequent in the framework of Tobit, and all of them happen to be direct objects, realizing the choice +Affected. This means that the sample estimate for the grammatical probability of +Affected is 100 selections/100 instances of the system. The typical spread around this value is 26.73 selections/100 instances of the system. This sample is skewed radically in a positive direction (3.33). This sample has the steepest peak of any I have observed, over four times as steep as the normal distribution (kurtosis of 12.08).

Thematic Complements are also infrequent in the framework of *Joseph and Aseneth*, and (like Tobit) all them are direct objects, realizing the choice +Affected. As a result, the sample estimate for the grammatical probability of +Affected is 100 selections/100 instances of the system. The typical spread around this value is 48.80 selections/100 instances of the system. This sample is strongly skewed in a positive direction (0.94). This sample has a flatter peak than the normal distribution (kurtosis of 1.9).

Thematic predicate nominatives and indirect objects are quite rare in all three samples. In fact, only Judith makes selections from subsystem 5 (+/-Stative) at all. The sample estimate for the grammatical probability of +Stative in Judith is 100 selections/100 instances of the system. The typical spread around this value is 35.56 selections/100 instances of the system. This sample is radically skewed in a positive direction (2.27). This sample has a peak twice as steep as that of the normal distribution (kurtosis of 6.15).

Quantifying the Projected Discourse of Non-Referential Narratives

Thematic verbs, realizing the choice +Process, are quite common in the projected discourse portions of Judith. The sample estimate for the grammatical probability of +Process is 66 selections/100 instances of the system. The typical spread around this value is 33.64 selections/100 instances of the system. This sample is negatively skewed (-0.48). It has a peak much flatter than the normal distribution (kurtosis of 1.27).

Thematic verbs are also quite common in the projected discourse portions of Tobit. The sample estimate for the grammatical probability of +Process is 73 selections/100 instances of the system. The typical spread around this value is 17.68

selections/100 instances of the system. This sample is strongly skewed in the negative direction (-1.30). This sample has a much steeper peak than the normal distribution (kurtosis of 5.19).

Thematic verbs are also common in the projected discourse portions of *Joseph and Aseneth*. The sample estimate for the grammatical probability of +Process is 62 selections/100 instances of the system. The typical spread around this value is 15.49 selections/100 instances of the system. This sample is heavily skewed in the positive direction (1.11). This sample is slightly more peaked than the normal distribution (kurtosis of 3.40).

Thematic adjuncts, realizing the choice +Circumstance, are fairly frequent in the projected discourse portions of Judith. The sample estimate for the grammatical probability of +Circumstance is 38 selections/100 instances of the system. The typical spread around this value is 23.41 selections/100 instances of the system. This sample is positively skewed (0.41). This sample has a peak much flatter than that of the normal distribution (kurtosis of 1.70).

Thematic adjuncts are fairly common in the projected discourse of Tobit. The sample estimate for the grammatical probability of +Circumstance is 33 selections/100 instances of the system. The typical spread around that value is 14.51 selections/100 instances of the system. This sample is strongly skewed in the negative direction (-1.13). This sample has a slightly steeper peak than the normal distribution (kurtosis of 3.16).

Thematic adjuncts are fairly common in the projected discourse portions of *Joseph and Aseneth*. The sample estimate for the grammatical probability of +Circumstance is 42 selections/100 instances of the system. The typical spread around this value is 26.86 selections/100 instances of the system. This sample is slightly skewed in the negative direction (-0.06). This sample has a much flatter peak than the normal distribution (kurtosis of 1.82).

Thematic subjects are fairly common in the projected discourse portions of Judith. The sample estimate for the grammatical probability for +Subject is 84 selections/100 instances of the system. The typical spread around this value is 46.30 selections/100 instances of the system. This sample is somewhat skewed in the negative direction (-0.38). It has a much flatter peak than the normal distribution (kurtosis of 1.31).

Thematic subjects are common in the projected discourse portions of Tobit. The sample estimate for the grammatical probability of +Subject is 50 selections/100 instances of the system. The typical spread around this value is 19.95 selections/100 instances of the system. This sample is substantially negatively skewed (-0.74). This sample has a steeper peak than that of the normal distribution (kurtosis of 3.60).

Thematic subjects are common in the projected discourse portions of *Joseph and Aseneth*. The sample estimate for the grammatical probability of +Subject is 87 selections/100 instances of the system. The typical spread around this value is 35.47 selections/100 instances of the system. This sample is skewed radically in the negative direction (-1.68). This sample has a somewhat steeper peak than the normal distribution (kurtosis of 4.40).

Thematic Complements are relatively infrequent in the projected discourse portions of Judith, but thematic direct objects, realizing the choice +Affected, are more common than either of the categories associated with subsystem 5 (+/- Stative). The sample estimate for the grammatical probability of +Affected is 60 selections/100 instances of the system. The typical spread around this value is 39.65 selections/100 instances of the system. This sample is heavily skewed in the positive direction (1.32). This sample has a slightly flatter peak than that of the normal distribution (kurtosis of 2.96).

Thematic Complements are relatively infrequent in the projected discourse portions of Tobit, but thematic direct objects, realizing the choice +Affected, are more common than either of the categories associated with subsystem 5. The sample estimate for the grammatical probability of +Affected is 57 selections/100 instances of the system. The typical spread around this value is 38.38 selections/100 instances of the system. The sample is slightly skewed in a negative direction (-0.10). It has a much flatter peak than the normal distribution (kurtosis of 1.58).

Thematic Complements are very infrequent in the projected discourse portions of *Joseph and Aseneth*; all four instances happen to select +Affected. This results in a sample estimate of the grammatical probability for +Affected of 100 selections/100 instances of the system. The standard deviation of this sample is 48.80 selections/100 instances of the system. This sample is strongly skewed in a positive direction (0.95). This sample has a flatter peak than the normal distribution (kurtosis of 1.9).

The sample from Judith contains only two instances of subsystem 5 (+/- Stative); both terms in the system appear once. This results in a sample estimate for the grammatical probability of +Stative of 50 selections/100 instances of the system. The typical spread around this value is 35.36 selections/100 instances of the system. This sample is radically skewed in the positive direction (2.27). The peak of this sample is a little over twice as steep as that of the normal distribution (kurtosis of 6.14).

The projected discourse portions of Tobit contain thirty-three instances of subsystem 5. The sample estimate for the grammatical probability of +Stative is 79 selections/100 instances of the system. The typical spread around this value is 43.10 selections/100 instances of the system. This sample is very slightly skewed in the negative direction, although it is symmetrical to the two decimal places I have been using. This sample has a much flatter peak than the normal distribution (kurtosis of 1.38).

The Validity of "Non-Referential Narrative" as a Genre Category

This section adds quantitative support to the qualitative assignment of Judith, Tobit, and *Joseph and Aseneth* to the same genre category, what I am terming "non-referential narrative." This quantitative support involves three tools: (1) the descriptive statistics reported in the preceding section, (2) confidence intervals for the relative frequencies on which the sample means (i.e. estimates of grammatical probability) are based, and (3) boxplots with confidence intervals for the median of the normalized frequencies. Tables 6.1 and 6.2 contain the confidence intervals of the relative frequencies for the framework and projected discourse, respectively; the boxplots appear as part of the explanations that follow.

Table 6.1: Confidence Intervals for the Relative Frequencies (Framework)

Subsystem	Judith	Tobit	*Joseph and Aseneth*
1 (+/- Process)	76±5.57	73±4.97	89±4.41
2 (+/- Circumstance)	37±12.88	45±10.70	41±20.55
3 (+/- Subject)	65±16.03	85±10.31	85±19.41
4 (+/- Affected)	92±15.35	100±0[1]	100±0
5 (+/- Stative)	100±0	0±0	0±0

[1] The confidence intervals with a range of 0 result from cases where the sample contained no realizations of the terms from the more delicate subsystem. See the discussion below.

Table 6.2: Confidence Intervals for the Relative Frequencies (Discourse)

Subsystem	Judith	Tobit	*Joseph and Aseneth*
1 (+/- Process)	66±7.46	52±4.55	62±8.13
2 (+/- Circumstance)	38±13.19	33±10.17	42±13.42
3 (+/- Subject)	84±12.70	50±8.00	87±12.78
4 (+/- Affected)	60±42.94	50±11.13	100±0
5 (+/- Stative)	50±69.29	79±13.90	0±0

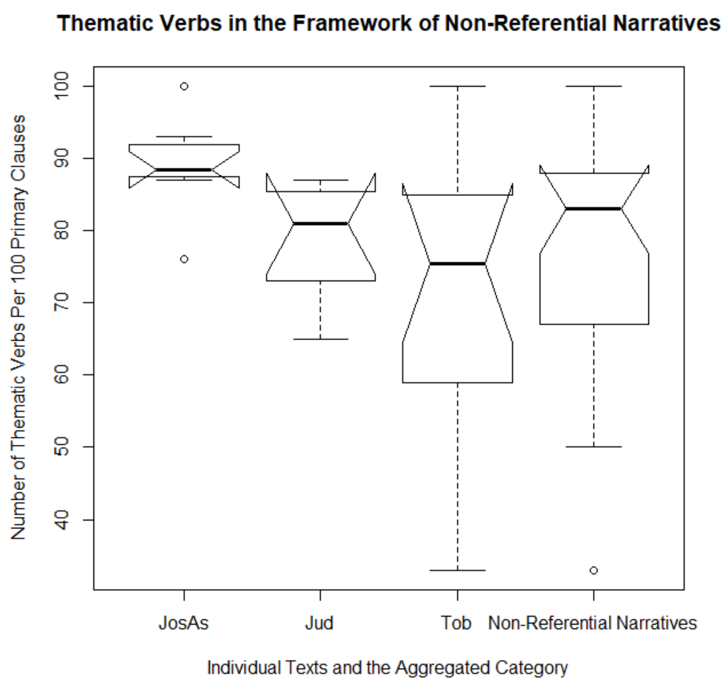
Do the Frameworks Point to A Coherent Category?

The basic descriptive statistics associated with the samples from Judith, Tobit, and *Joseph and Aseneth* on the use of subsystem 1 (+/- Process) paint an interesting picture of the potential genre relationship between them. Although the standard deviation of the sample from Tobit diverges substantially from the other two, the standard deviations of the samples from Judith and *Joseph and Aseneth* differ by less than two selections/100 instances of the system. All three samples exhibit significant negative skew; the magnitude of said skew is almost precisely the same for the samples from Tobit and *Joseph and Aseneth*. As for kurtosis, the samples from Judith and Tobit have flatter peaks than the normal distribution, while the sample from *Joseph and Aseneth* has a steeper peak than the normal distribution. Thus, only the skewness suggests that all three texts belong together, but the standard deviation and kurtosis disagree with regard to which two belong together. Given that these statistics are calculated from the normalized frequencies, some fluctuation is probably inevitable; nevertheless, these data present limited support for combining these three samples.[2]

The confidence intervals for the grammatical probability of +Process based on relative frequencies strongly support classifying Judith and Tobit as belonging to the

[2] See “Lessons Learned” in Chapter 8.

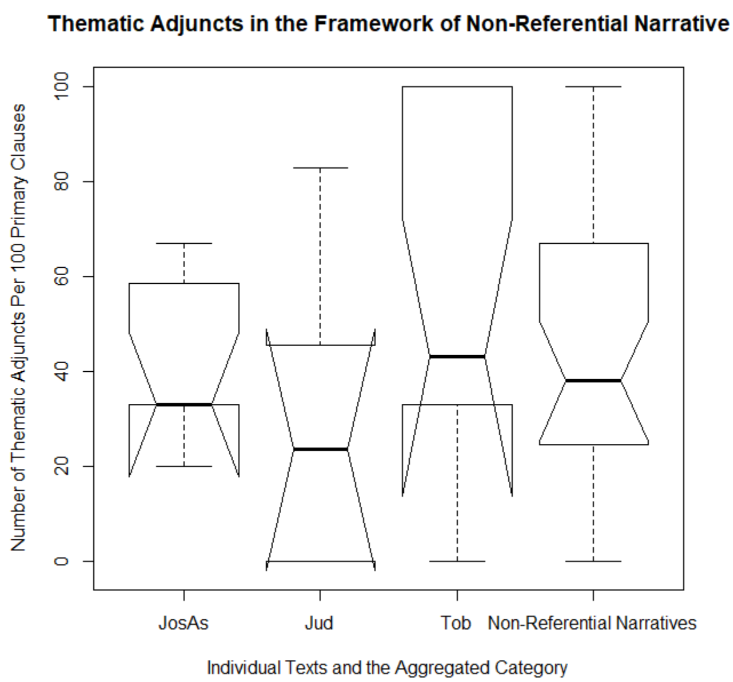
same genre, but not *Joseph and Aseneth*. The confidence intervals for the first two samples far surpass the minimum standard of overlapping: both sample means lie within the confidence interval for the other mean. A gap of roughly 3 selections/100 instances of the system separates the confidence intervals for the samples from Judith and *Joseph and Aseneth*; the gap between the confidence intervals for the samples from Tobit and *Joseph and Aseneth* lies between 6 and 7 selections/100 instances of the system. Thus, the data as they currently stand do not support including *Joseph and Aseneth*. The gap is small enough, however, that increasing the size of the sample easily could fix the problem, as the boxplot below illustrates.



This boxplot loosely supports treating all three texts as instances of a single genre. As the data stand, they barely meet the minimum requirement: all three confidence intervals overlap, albeit by less than 1 selection/100 instances of the system. The real margin of overlap is likely to be higher than this. The shape of the box for the sample from *Joseph and Aseneth* raises concerns regarding the representativeness of this sample. Notice that only the top half of the box has the trapezoidal shape a notched boxplot should have; the bold line marking the sample median is so close to the bottom of the box that the confidence interval protrudes beyond the box, even though the confidence interval is quite narrow. Increasing the length of the sample will almost certainly remedy this problem one way or the other.

If the additional data trends above the current sample median, the value of the sample median will increase but the data below the current median, by definition 50 percent of the currently collected data, will still be there, dragging the sample median away from the bottom of the box. If the additional data trends below the current sample median, the median will come down, but the data below the current median will still be there, pulling the bottom of the box away from the median. In the former case, the median will be moving away from the other two samples and, as a result, the margin of overlap between their confidence intervals would decrease, given a consistent width; in the latter case, the median will be moving toward the medians of the other two samples and, as a result, the margin of overlap between their confidence intervals will increase, given a consistent width. Since the current value for the median is between 88 and 89 selections/100 instances of the system, a decrease seems more likely than an increase. By the same logic, assuming the normalization process affected all three samples equally so that I can directly compare the confidence intervals in the boxplot and the ones calculated from the relative frequencies, the confidence interval for the grammatical probability of +Process in the sample from *Joseph and Aseneth* should move closer to the ones for the other two samples as well. As a result, I conclude that, on balance, the framework data on subsystem 1 (+/- Process) is consistent with all three samples belonging to the same genre.

The data on the use of subsystem 2 (+/- Circumstance) strongly support considering them as a single category. Of the basic descriptive statistics, only the skewness of the sample from *Joseph and Aseneth* is out of line: it shows a slight negative skew, whereas the other samples are positively skewed. On the other hand, the largest difference in standard deviations is less than 8 selections/100 instances of the system. All three samples have peaks flatter than the normal distribution. The confidence intervals for the grammatical probability of +Circumstance calculated from the relative frequencies overlap; moreover, all three of the sample means fall within the narrowest confidence interval, providing clear evidence that these data form a coherent group. The boxplot for these data is below.

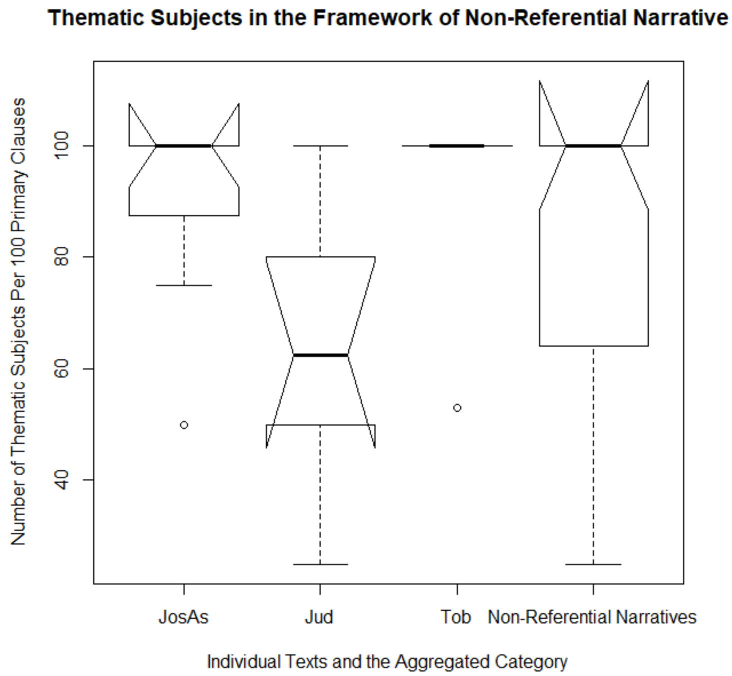


This boxplot presents a somewhat unclear picture of the samples' relationship to one another, for each box associated with a sample has at least one set of protruding triangles that designate a problematic confidence interval boundary: the lower bound of the interval for *Joseph and Aseneth* passes the box by roughly 15 selections/100 instances of the system, the lower bound of the interval for Judith passes the box by roughly 2 selections/100 instances of the system, the upper bound of the interval for Judith passes the box by between 3 and 4 selections/100 instances of the system, and the lower bound of the interval for Tobit passes the box by roughly 20 selections/100 instances of the system. These problematic boundaries, particularly the ones for *Joseph and Aseneth* and Tobit, greatly impair the usefulness of the confidence intervals for the sample median, represented by the notches in the plot, because the estimates behind these intervals are not sufficiently precise for overlap between their intervals to be particularly meaningful. In other words, although the confidence intervals for all three samples overlap by over 30 selections/100 instances of the system, the individual confidence intervals are wide enough that at least some of that overlap results from the sample estimates for the population median being too imprecise, rather than being the real overlap that would indicate the texts in question come from a coherent group.

Nevertheless, particularly in light of the basic descriptive statistics and the confidence intervals for the relative frequencies having supported the validity of the category, the box for the aggregated data suggests that the texts from which the samples come are more similar than the samples themselves indicate, i.e. that the problem is one of sample representativeness. Neither boundary for the confidence interval of the aggregated data protrudes beyond the box; even the lower boundary, which at the scale necessary to plot the individual samples seems to meet the edge of the box, lies inside it by almost 4 selections/100 instances of the system. If the data did not belong to the same underlying distribution, combining them would produce a messier plot, not a clearer one. Thus, in summary, I conclude that the boxplot for these data offer limited support for the coherence of these samples, though additional work should be done, in terms of lengthening these samples, adding new samples, improving the normalization procedure, or all of the above.[3]

Turning to subsystem 3 (+/- Subject), the basic descriptive statistics offer only weak support to the hypothesis that the three texts under discussion come from the same genre. The standard deviation for the sample from Tobit is almost exactly double that of the sample from Judith with the sample from *Joseph and Aseneth* falling in the middle. All three samples are negatively skewed, but the magnitude of said skew varies widely. The samples from Judith and Tobit have peaks flatter than the normal distribution, while the sample from *Joseph and Aseneth* has a steeper peak. The confidence intervals for the grammatical probability of +Subject based on relative frequencies overlap substantially: the interval for *Joseph and Aseneth* completely subsumes the interval for Tobit, and the sample mean for Judith lies less than 1 selection/100 instances of the system outside the interval for *Joseph and Aseneth*. The boxplot for this group of data is below.

[3] See "Lessons Learned" in Chapter 8.

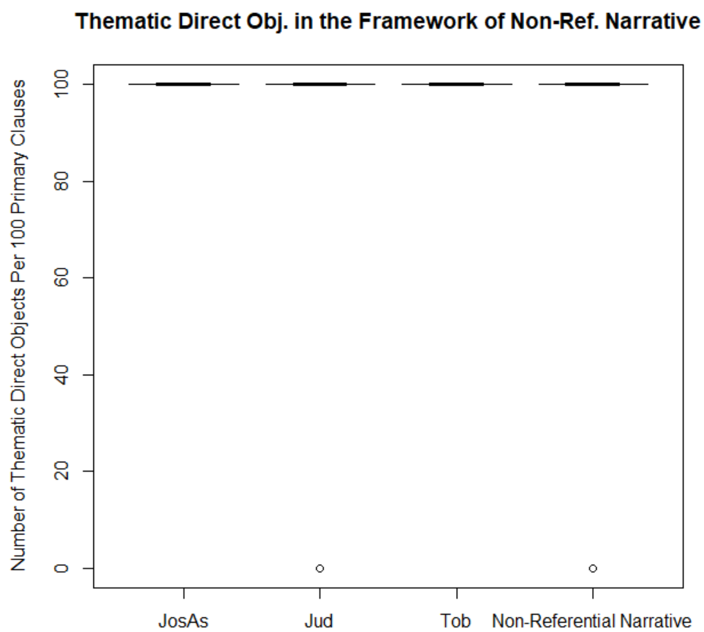


This boxplot does not support treating these three texts as representative of a single genre. Like the previous boxplot representing subsystem 2 (+/- Circumstance), all three boxes associated with samples have at least one problematic confidence interval boundary; unlike that plot, however, the box for the aggregated data is also problematic here. As these data stand, the confidence intervals for the sample medians do not overlap, suggesting a statistically significant difference between the population medians, if the sample estimates for the median are accurate. This plot conclusively demonstrates, however, that the sample estimates are not accurate. If the medias of the samples from *Joseph and Aseneth*, Tobit, and the aggregated data were reliable estimates of the median in their respective populations, two patently absurd conclusions would follow. First, since confidence intervals for a median extend on both sides by definition, a median of 100 selections/100 instances of the system would imply that some texts in the population had a median above that figure and, thus, selected +Subject more times than the choice was available. Second, thematic Complements would not exist, for there would be no instances of -Subject to open subsystems 4 and 5. Therefore, the sample estimates cannot be reliable. Significantly, however, this does not mean that the plot supports treating the data as a unified genre category; it only shows that I have not conclusively disproven that they belong to one. Clearer light on the question awaits additional data and especially improvements to the normalization procedure on which the boxplots are based.[4] For the time being, the other tools showed that the data are at least consistent with having come from a single genre, so problems with the boxplot do not seem an insurmountable hurdle.

Thematic Complements, the categories associated with subsystem 4 (+/- Affected) are sufficiently infrequent within the framework portions of the three texts I sampled (between 2 and 11 instances, depending on the sample) that ascertaining the relationship among them is difficult. For the sake of completeness, however, I include a discussion here. The larger gap in standard deviations, between Tobit and *Joseph and Aseneth*, is 22.07 selections/100 instances of the system. These two samples are both radically skewed in the positive direction, but the sample from Judith is quite skewed in the negative direction. The sample from Judith has a peak flatter than the normal distribution, but the other two samples have a peak much steeper than the normal distribution. The similarities between the samples from Tobit and *Joseph and Aseneth* seem to result from both samples happening to contain no instances of -Affected. This similarity also manifests in the confidence intervals for the grammatical probability of +Affected using relative frequencies: no instances of -Affected in the sample means that the only conclusion to draw on the basis of that sample is that there is no variation in the system.

[4] See “Lessons Learned” in Chapter 8.

Clearly, however, predicate nominatives and indirect objects occur in thematic position within other texts, so this would seem to be an artifact of having very few instances, unless one would like to make the case that these particular populations never place predicate nominatives and/or indirect objects in thematic position. The boxplot for these data is below.



This plot illustrates even more graphically the problems observed with the preceding plots. The confidence intervals overlap, but only because the middle 50 percent of the data for each sample congregates at the upper extreme, a theoretically improbable case. In summary, this plot indicates the need for a better normalization procedure; the categories involved are infrequent enough that a brute force solution via the addition of more data seems unlikely to succeed.[5]

[5] See “Lessons Learned” in Chapter 8.

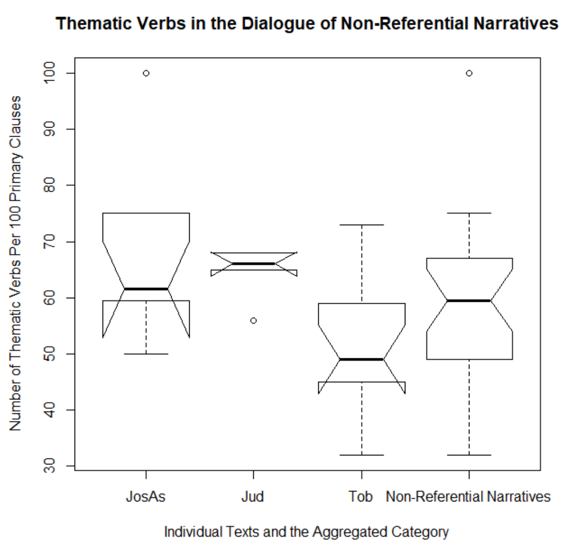
The above constitutes all available comparisons to ascertain whether the quantitative data point towards the framework portions of Judith, Tobit, and *Joseph and Aseneth* representing a coherent genre category. I performed three comparisons (basic descriptive statistics, confidence intervals based on relative frequencies, and boxplots with confidence intervals for the median normalized frequency) for each subsystem, but the problems I have observed with the normalization procedure, which lies behind both the basic descriptive statistics and the boxplots, indicate that the confidence intervals based on relative frequencies bear special weight.

The comparisons for subsystem 1 (+/- Process) strongly pointed towards Judith and Tobit belonging to the same genre, but the relationship of the sample from *Joseph and Aseneth* to the others was less clear, though I think much of that could be remedied by lengthening that sample. All the comparisons of the data on subsystem 2 (+/- Circumstance) supported the genre coherence of these three samples, though the boxplot raised some concerns of sample representativeness that could be shored up with additional data and/or procedural improvements. Only one comparison of the data on subsystem 3 (+/- Subject) portrayed these three samples as sharing a genre. On the other hand, the comparison in question is the most reliable one, the confidence intervals based on the relative frequencies, and I found clear reasons to suspect the normalization procedure created the problems with the other two comparisons. None of the comparisons of the data on subsystem 4 (+/- Affected) clearly placed all three samples together with regards to genre, although the samples from Tobit and *Joseph and Aseneth* stood out together for having no instances of -Affected, which incidentally prevented in any comparisons for subsystem 5 (+/- Stative) because only Judith contained any selections from it. These comparisons collectively present a substantial case that the quantitative data regarding clause thematization in the frameworks of Judith, Tobit, and *Joseph and Aseneth* are at least consistent with all three belonging to the same genre, and in some cases point substantially in that direction.

Do the Projected Discourse Portions Point to a Coherent Category?

The basic descriptive statistics on the use of subsystem 1 (+/- Process) do not support combining these three samples. The standard deviation of the sample from Judith is over twice that of the sample from *Joseph and Aseneth*. The sample from Judith is moderately skewed in the negative direction, the sample from Tobit is heavily skewed in the negative direction, and the sample from *Joseph and Aseneth* is heavily skewed in the positive direction. The sample from Judith has a much flatter peak than the normal distribution, the sample from Tobit has a much steeper peak than the normal distribution, and the sample from *Joseph and Aseneth* has a peak somewhat steeper than the normal distribution but not nearly as steep as the peak of the sample from Tobit.

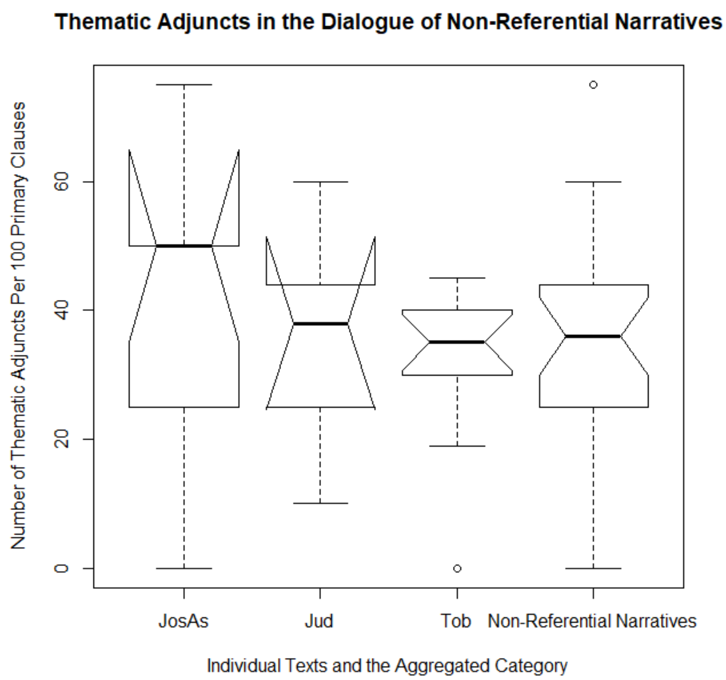
As the data currently stand, the confidence intervals based on relative frequencies for the grammatical probability of +Process present a somewhat convoluted picture of the potential genre relationship among the three texts sampled. The narrowest of the three intervals (the one for Tobit) overlaps with only the widest (the one for *Joseph and Aseneth*); a gap of almost exactly 2 selections/100 instances of the system remains between the upper bound of the confidence interval for Tobit and the lower bound of the one for Judith. On the other hand, the confidence intervals for Judith and *Joseph and Aseneth* are strongly consistent with each other, for both sample estimates for the grammatical probability of +Process fall well within the other interval. These three observations (the gap between Judith and Tobit, the substantial overlap between Judith and *Joseph and Aseneth*, and the marginal overlap between Tobit and *Joseph and Aseneth*) point in one of two directions: either the samples are in fact consistent with each other and the sample from Tobit and the sample from Judith are slightly skewed relative to their populations, i.e. the observed gap of 2 selections/100 instances of the system between the confidence intervals results from sampling error, or the apparent overlap between Tobit and *Joseph and Aseneth* is a mirage resulting from the standard error of the sample proportion from *Joseph and Aseneth* being too high, i.e. too much uncertainty surrounds the characteristics of *Joseph and Aseneth* and, therefore, the confidence interval for *Jospeh and Aseneth* includes values it should not. Collecting more data would be the only definitive way to adjudicate which of these scenarios is actually the case; for the time being, I am forced to conclude that the confidence intervals based on the relative frequencies of +Process and -Process point to Judith and *Joseph and Aseneth* sharing a genre, but they do not support including Tobit within it. A boxplot for these data is below.



In general terms, this boxplot tells the same story as the confidence intervals discussed in the previous paragraph: the confidence interval for the median of the sample from *Joseph and Aseneth* overlaps with the confidence intervals associated with the other two texts, but those confidence intervals do not overlap with each other. The plot also obviously shows problems with the representativeness of all three samples; interestingly, the problems all concern the lower bound of their respective distributions. The box for *Joseph and Aseneth* shows a low degree of confidence in the sample median as an estimator for the population median, as the confidence interval for the median passes the box by roughly 6 selections/100 instances of the system. In and of itself, this observation would indicate that the marginal overlap between the confidence intervals based on relative frequencies, discussed in the previous paragraph, was a mirage. Two other factors, however, counterbalance this. First, the plot also shows that the median of the sample from Judith is also a poor estimate for its population median, and the reason for this is that the sample median is very close to the bottom of the box. Thus, it is debatable whether the problem lies with the confidence interval for *Joseph and Aseneth* being problematically wide or the confidence interval for Judith being too narrow. Second, simply combining all three samples produces an almost ideal-looking plot, which seems more likely in the case of three samples from the same distribution than in the case of three samples of different distributions. In summary, the data on subsystem 1 (+/- Process) seem at least somewhat consistent with all three samples belonging to the same genre, though additional work needs to be done to strengthen this likelihood.

The basic descriptive statistics on the use of subsystem 2 (+/- Circumstance) within the projected discourse portions of the three samples do not support their genre coherence. The larger difference in standard deviations is a little over 12 selections/100 instances of the system. The sample from Judith is skewed in the positive direction, while the other two are skewed in the negative direction. The sample from Tobit has a steeper peak than the normal distribution, but the other two have a flatter peak.

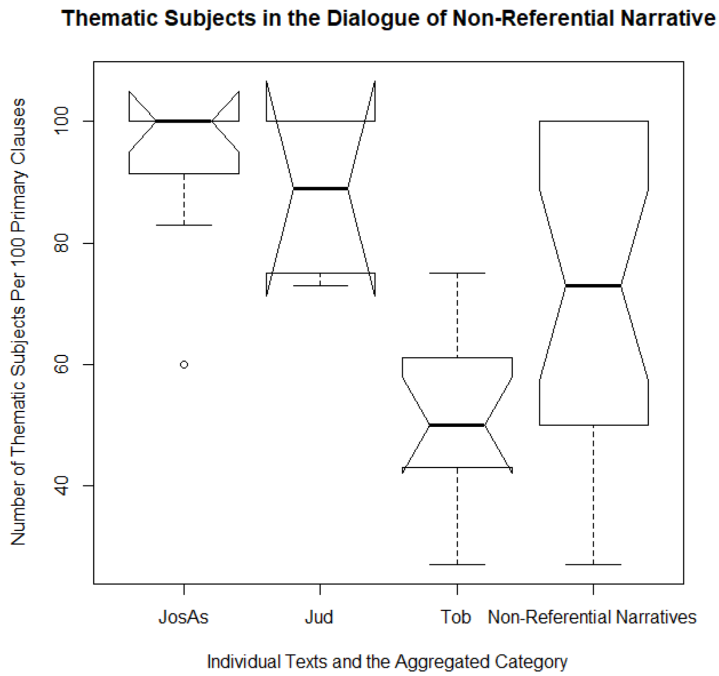
The confidence intervals based on relative frequencies strongly suggest that all three texts belong to the same genre. The region shared between all three confidence intervals measures 14.5 selections/100 instances of the system. Moreover, the narrowest confidence interval contains all three sample estimates. A boxplot of these data is below.



As these data stand, the confidence intervals for the median overlap, although the accuracy of the confidence intervals for the samples from *Joseph and Aseneth* and Judith is questionable, as the protruding triangles indicate. Significantly, however, the problem in both cases is on the upper end of the distribution, while the overlap with Tobit's confidence interval is entirely below the median in the case of *Joseph and Aseneth* and partially below the median in the case of Judith. While lengthening the samples so that they will better estimate their respective population medians would still be worthwhile, the observation that the well-formed parts of the distribution still overlap ameliorates concerns regarding a false positive. In particular, the position of the current apparent overlap suggests the problem lies with the representativeness of the sample from *Joseph and Aseneth*, especially since even the problematically wide confidence interval for Judith barely extends to the median of the sample from *Joseph and Aseneth*. The box for the aggregated data reinforces this conclusion: its confidence interval falls well within the box, and its median includes the sample medians for Judith and Tobit. Overall, the data are compatible with the three texts having come from the same genre, and some of the indicators move beyond this to substantially supporting that they share a genre.

Turning to subsystem 3 (+/- Subject), the basic descriptive statistics are divided. On the one hand, all three samples are negatively skewed, which is at least consistent with all three samples having come from the same population. On the other hand, the standard deviation of the sample from Tobit is far less than the standard deviations of the other two samples, and the sample from Judith has a flatter peak than the normal distribution, unlike the other two samples.

Likewise, all three confidence intervals based on relative frequencies do not overlap. The confidence intervals associated with the samples from Judith and *Joseph and Aseneth* overlap substantially because they are almost equally wide and their centers are only 3 selections/100 instances of the system apart. The upper bound of the confidence interval for the sample from Tobit lies over 13 selections/100 instances of the system below the lower bound of the confidence interval associated with Judith. Thus, the confidence intervals based on relative frequencies strongly support there being a common distributional pattern behind the data from Judith and *Joseph and Aseneth*, but Tobit probably does not share that pattern. The boxplot for these data is below.



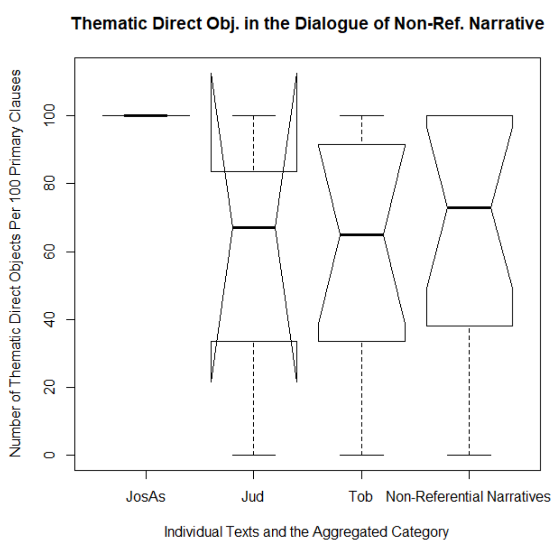
The plots for all three samples have the triangles that indicate the confidence interval for the sample's median extends beyond the box, which normally signals the need for a longer sample. In this case, the upper bounds of the confidence interval for two of the sample medians, namely the ones for *Joseph and Aseneth* and Judith, also extend into theoretically impossible values (roughly 105 and 107 selections/100 instances of the system). Furthermore, the lower bound of the confidence interval for the median of the Judith sample is also problematic, extending below the box by roughly 4 selections/100 instances of the system. Lastly, the lower bound of the confidence interval for Tobit's sample median extends beyond the box by less than 1 selection/100 instances of the system.

The data as they stand do not support combining *Joseph and Aseneth*, Judith, and Tobit into one genre category. While the confidence intervals for the sample medians of the samples from *Joseph and Aseneth* and Judith overlap, supporting the intuition that they belong to the same genre, neither of them overlap with the confidence interval for the median of the sample from Tobit. In fact, the value of 60 selections/100 instances of the system, which qualifies as a mathematical outlier in the sample of *Joseph and Aseneth* (as the open dot in the plot indicates), still lies above the upper bound of the confidence interval for the median of the Tobit sample by a little more than 2 selections/instances of the system.

The plot of the aggregated data as they stand has the shape one would expect of a single, wide distribution: the confidence interval for the median is well within the box, and the data point that was an outlier on the individual sample level is now not even outside the range of likely values for the median. Nevertheless, both the problems discussed above and the width of the confidence interval (nearly 32 selections/100 instances of the system) suggest a low degree of confidence in the median value of 73 selections/100 instances of the system as a reliable estimate of the grammatical probability in the genre as a whole, if in fact these data come from a single genre at all. Additional samples would be the means of making the estimate more reliable.

Turning to subsystem 4 (+/- Affected), the basic descriptive statistics are, for the most part, consistent with the samples having come from the same population. The gap between the largest and smallest standard deviations is a little more than 10 selections/100 instances of the system. All three samples have a flatter peak than the normal distribution, although the kurtosis of the sample from Judith is closer to that of the kurtosis of the normal distribution than to the kurtosis of the other two samples. The measure that disagrees is skewness: the sample from Tobit exhibits slight negative skew, while the other two are substantially skewed in the positive direction. Overall, however, it would be fair to say that this indicator offers limited support to the compatibility of these three samples.

Assessing the compatibility of the three samples using the confidence intervals based on relative frequencies is difficult in this case because the projected discourse portions of *Joseph and Aseneth* contain no instances of -Affected. As a result, the standard error of the proportion comes out to be 0, and multiplying 0 by 1.96 still gives 0 as the width of the confidence interval. In other words, there is no confidence *interval* for the data on subsystem 4 (+/- Affected), there is only the point estimate. Nevertheless, the narrower of the two remaining confidence intervals contains the sample estimates for the grammatical probability of +Affected in both Judith and Tobit; the wider confidence interval contains the sample estimate from *Joseph and Aseneth* as well, though it is admittedly quite wide. The boxplot for these data is below.



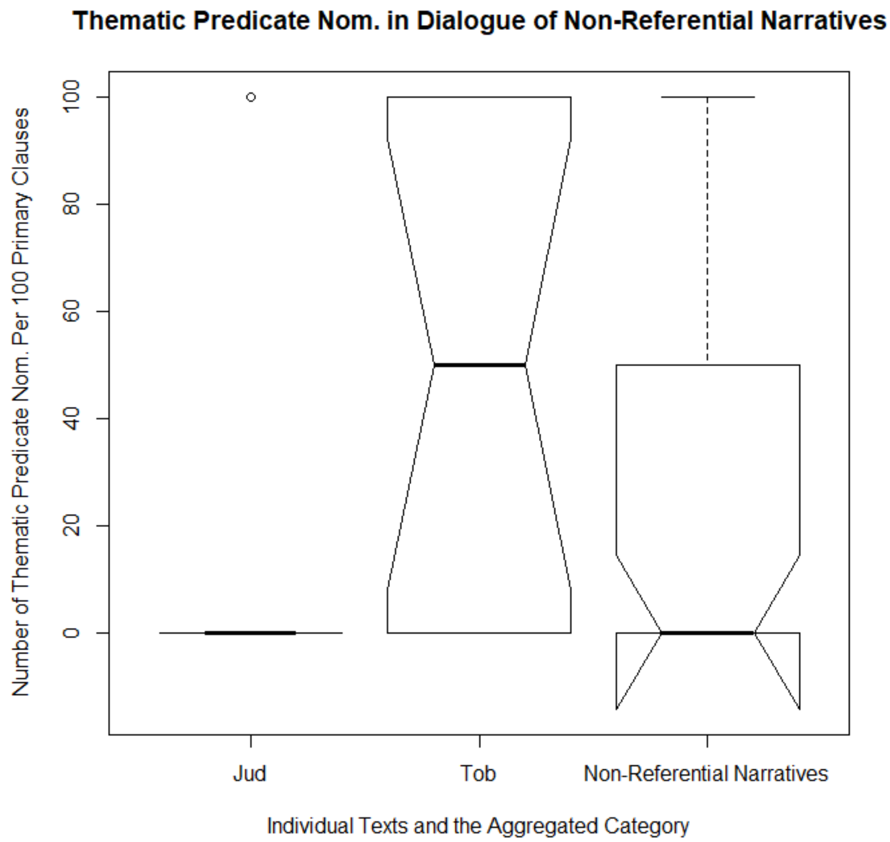
The samples from *Joseph and Aseneth* and Judith both exhibit problems with the confidence intervals for their medians. The problem with the sample from *Joseph and Aseneth* is the small (actually, in this case, nonexistent) range of the data; all the data points are at 100 selections per 100 instances of the system. As I have pointed out several times now, logically this extreme result cannot be a reliable estimate for any population, and it points to problems with my data collection procedure. In the case of the sample from Judith, the confidence interval for the median is too wide, spanning roughly 92 selections/100 instances of the system, the top 13 of which are theoretically impossible values. Extending the samples may help with this, but the crux of the problem is the normalization procedure I followed.

Asking whether the medians are compatible with each other and, thus, support combining the samples into one category is almost meaningless when two of the confidence intervals are as problematic as those for the medians of the samples from *Joseph and Aseneth* and Judith. Any median whose median included any numbers between 21 selections/100 instances of the system and 100 selections/100 instances of the system would be compatible with the sample from Judith; this includes both *Joseph and Aseneth* and Tobit (although, notably, the latter two do not overlap). While extending the samples I already have and taking additional ones could help clarify this question, neither is likely to be effective while the current normalization procedure remains in place.

The plot of the aggregated data looks more or less like the plot one would expect to see for a single, widely disbursed distribution, although it is noticeably skewed towards the higher values. The problems described above and the extreme width of the confidence interval (roughly 48 selections/100 instances of the system) both indicate a low degree of confidence in the median as a reliable estimate of the true grammatical probability in the genre as a whole (assuming the three samples in fact represent a single genre). Additional data, either in the form of extending the samples I currently have, integrating new ones, or (ideally) both, should help with this problem, but issues with the normalization procedure should be fixed first in order to maximize the effect of the new data.

The projected discourse portions of *Joseph and Aseneth* make no selections from subsystem 5 (+/- Stative), and the sample from Judith contains only two, but a comparison of Judith and Tobit follows for the sake of completeness. The standard deviations of these two samples differ by less than 8 selections/100 instances of the system. The skewness and kurtosis of the samples do not suggest that the samples belong together: the sample from Judith strongly skewed in the positive direction with a peak much steeper than the normal distribution, and the sample from Tobit is very slightly skewed in the negative direction with a peak flatter than the normal distribution.

The confidence intervals based on relative frequencies for the samples from Judith and Tobit overlap, but Judith’s interval is wide enough that it covers the entire range of theoretically possible values. This observation, combined with the narrower confidence interval not containing the mean of the broader one, leads me to conclude that these samples probably do not result from the same distributional pattern.



This boxplot clearly suffers from the small sample size and problems with the normalization procedure.[6] Problems with the normalization procedure mean that the one instance of +Stative shows up as a mathematical outlier: the one instance of -Stative is indistinguishable from the bulk of the chapters where there are no choices from this subsystem, for both are recorded as 0 selections/100 instances of the system. Likewise, although Tobit contains 33 choices from this subsystem with both terms being represented, a number of these wound up combined with others because they tended to occur in clumps within the same chapter.

[6] See “Lessons Learned” in Chapter 8.

In summary, the use of subsystem 1 (+/- Process) in the projected discourse portions of each sample does not strongly support classifying Judith, Tobit, and *Joseph and Aseneth* together, although the boxplot in particular suggests that they might do so if the samples from Judith and *Joseph and Aseneth* are lengthened. Even as they stand, however, the samples are at least broadly consistent with having come from the same genre. The use of subsystem 2 (+/- Circumstance) in the projected discourse portions of each sample is consistent with their texts belonging to the same genre, and some of the indicators moved beyond this minimum bar and provided substantial evidence that the texts do belong to the same genre. The use of subsystem 3 (+/- Subject) in the projected discourse portions of each sample is consistent with Judith and *Joseph and Aseneth* belonging to the same genre, but they suggest that Tobit might not belong with them, at least on the basis of thematization patterns. The projected discourse portions of these three texts make too few selections from subsystem 4 (+/- Affected) and subsystem 5 (+/- Stative) to form a coherent picture of how they use these subsystems. Although in all cases lengthening the samples from Judith and *Joseph and Aseneth* as well as adding samples would be helpful, only in the case of thematic subjects (subsystem 3) do the current data seem insufficient to establish at least the plausibility of a genre relationship among the three texts.

Summary: The Overall Validity of "Non-Referential Narrative"

This subsection summarizes and integrates the discussion in the previous two subsections. The subsystem-by-subsystem comparisons of which those subsections consisted reached one of three conclusions regarding the relationship between each corresponding group of quantitative data and the *a priori* hypothesis, initially reached on qualitative grounds, that these samples come from texts belonging to the same population (i.e. genre): (1) the quantitative data are inconsistent with the qualitative hypothesis, (2) the quantitative data are consistent with the qualitative hypothesis but not strong enough to lend support to it, or (3) the quantitative data lend substantial support to the qualitative hypothesis.

Four of the seven comparisons performable for all three texts fell into the category of being consistent with the qualitative hypothesis previously adopted but not substantially supporting it. The predominance of this category is to be expected, given the low bar for finding data to be consistent, namely that I failed to prove the difference between them was statistically significant, and the high bar I set myself for claiming that the data quantitatively supported my prior qualitative hypothesis. Only one group of comparisons, the ones regarding the use of subsystem 2 (+/- Circumstance) in the framework, fell as a whole into the category of actively supporting the hypothesis, though individual indicators fell into this category on several other occasions. This leaves two groups of comparisons in the inconsistent category, namely the comparisons for subsystem 4 (+/- Affected) in the framework and subsystem 3 (+/- Subject) in projected discourse.

These results, if weighted equally, would suggest that the category of non-referential narrative was invalid. They should not be weighted equally, however, because the reliability of the comparisons depends on the reliability of the data upon which they are based. The reliability of each group of data depends in turn on the amount of information that went into it (essentially the sample size). The comparisons within groups already take this into consideration because the amount of information that went into the calculation of a sample median or proportion is precisely what the standard error portion of the formula for the width of a confidence interval is measuring, so comparing overlap between confidence intervals amounts to asking whether the divergence between sample estimates within a group is greater than would be expected, given the standard error, i.e. the reliability, of each estimate. What is needed here is to take the same principle and apply it to comparisons between groups. This involves comparing raw standard errors, instead of confidence intervals.

Even without calculations, however, it is obvious that the group of comparisons that substantially supported treating these three samples together (the comparisons with regard to use of subsystem 2 in the framework) is based on a larger sample size (and, thus, should prove more reliable) than the first category indicating inconsistency among them, namely the comparisons for subsystem 4 (+/- Affected), because the latter is part of the former. The degree of imbalance is striking: the smallest number of instances of subsystem 2 (+/- Circumstance) in the framework of one of these samples is 22 in the framework of *Joseph and Aseneth*, whereas thee total sample size for subsystem 4 (+/- Affected), across the three samples is 21.[7] In the other case, the standard error for the proportion of +Circumstance to -Circumstance in the frameworks of these samples is 3.91 selections/100 instances of the system, whereas the standard error for the proportion of +Subject to -Subject in the projected discourse of these samples is 3.36 selections/100 instances of the system. This indicates that, as the data currently stand, the inconsistent group of comparisons should receive slightly more weight. Given that the difference is roughly half of one selection/100 instances of the system and the presence of individual

[7] As another indicator of poor sample representativeness, neither the framework of Tobit nor the framework of *Joseph and Aseneth* contain any instances of -Affected.

indicators providing substantial support in cases where I concluded that the entire group of comparisons was only consistent with the prior hypothesis, however, the data still seem, at worst, consistent with the prior hypothesis that the three texts in question share a genre.[8]

Thematization in Non-Referential Narrative: Overview[9]

Table 6.3 Overview of Thematization in Judith

	Circumstances	Process	Subject	Complement	Total
Narrative	20	172	22	12	226
Projected	20	103	27	5	155
Total	40	275	49	17	381

Table 6.4 Overview of Thematization in Tobit

	Circumstances	Process	Subject	Complement	Total
Narrative	37	224	39	7	307
Projected	74	240	75	75	464
Total	111	464	114	84	771

Table 6.5 Overview of Thematization in Joseph and Aseneth

	Circumstances	Process	Subject	Complement	Total
Narrative	9	171	11	2	193
Projected	22	85	26	4	137
Total	31	256	37	6	330

[8] The wild card in this discussion is the middle category: the validity of "non-referential narrative" would be easier to assess if each group of data presented a clear piece of evidence, either in favor or against, rather than only indicating that I could not disprove that the samples in question were reasonably similar. Lengthening the samples, particularly the one from *Joseph and Aseneth*, would help with this. Doing so could also help the comparisons that were already clear become even more so: inspecting the standard error associated with the group of comparisons that strongly indicated that the three texts shared a genre at the level of individual samples reveals that the standard error of the sample from *Joseph and Aseneth* is roughly twice that of the sample from Tobit.

[9] Tables 6.3–13 fit the same mold as the corresponding ones for Chapter 5, so I will not explain them again.

Thematization in Non-Referential Narrative: Framework

Table 6.6 Frequencies of Thematization Choices in Judith (Framework)

Subsystem	Plus Feature	Minus Feature	Total
1 (+/- Process)	172 (0.76)	54 (0.24)	226
2 (+/- Circumstance)	20 (0.37)	34 (0.63)	54
3 (+/- Subject)	22 (0.65)	12 (0.35)	34
4 (+/- Affected)	11 (0.92)	1* (0.08)	12
5 (+/- Stative)	1* (1.00)	0* (0.00)	0

Table 6.7 Frequencies of Thematization Choices in Tobit (Framework)

Subsystem	Plus Feature	Minus Feature	Total
1 (+/- Process)	224 (0.73)	83 (0.27)	307
2 (+/- Circumstance)	37 (0.45)	46 (0.55)	83
3 (+/- Subject)	39 (0.85)	7 (0.15)	46
4 (+/- Affected)	7 (1.00)	0* (0.00)	7
5 (+/- Stative)	0* (0.00)	0* (0.00)	0

Table 6.8 Frequencies of Thematization Choices in Joseph and Aseneth (Framework)

Subsystem	Plus Feature	Minus Feature	Total
1 (+/- Process)	171 (0.89)	22 (0.11)	193
2 (+/- Circumstance)	9 (0.41)	13 (0.59)	22
3 (+/- Subject)	11 (0.85)	2* (0.15)	13
4 (+/- Affected)	2* (1.00)	0* (0.00)	2
5 (+/- Stative)	0* (0.00)	0* (0.00)	0

Table 6.9 Normed Aggregate Frequencies (Framework)

Subsystem	Plus Feature (Per 100 Clauses)	Minus Feature (Per 100 Clauses)
1 (+/- Process)	79	21
2 (+/- Circumstance)	41	59
3 (+/- Subject)	78	22
4 (+/-Affected)	100	0
5 (+/-Stative)	100	0

Thematization in Non-Referential Narrative: Discourse

Table 6.10 Frequencies of Thematization Choices in Judith (Discourse)

Subsystem	Plus Feature	Minus Feature	Total
1 (+/- Process)	103 (0.66)	52 (0.34)	155
2 (+/- Circumstance)	20 (0.38)	32 (0.62)	52
3 (+/- Subject)	27 (0.84)	5 (0.16)	32
4 (+/- Affected)	3* (0.6)	2* (0.4)	5
5 (+/- Stative)	1* (0.5)	1* (0.5)	2

Table 6.11 Frequencies of Thematization Choices in Tobit (Discourse)

Subsystem	Plus Feature	Minus Feature	Total
1 (+/- Process)	240 (0.52)	224 (0.48)	464
2 (+/- Circumstance)	74 (0.33)	150 (0.67)	224
3 (+/-Subject)	75 (0.50)	75 (0.50)	150
4 (+/-Affected)	43 (0.57)	33 (0.43)	76
5 (+/- Stative)	26 (0.79)	7 (0.21)	33

Table 6.12 Frequencies of Thematization Choices in Joseph and Aseneth (Discourse)

Subsystem	Plus Feature	Minus Feature	Total
1 (+/- Process)	85 (0.62)	52 (0.38)	137
2 (+/- Circumstance)	22 (0.42)	30 (0.58)	52
3 (+/- Subject)	26 (0.87)	4 (0.13)	30
4 (+/- Affected)	4 (1.00)	0 (0.00)	4
5 (+/- Stative)	0 (0.00)	0 (0.00)	0

Table 6.13 Normed Aggregate Frequencies (Discourse)

Subsystem	Plus Feature (Per 100 Clauses)	Minus Feature (Per 100 Clauses)
1 (+/- Process)	60	40
2 (+/- Circumstance)	38	62
3 (+/- Subject)	86	14
4 (+/- Affected)	74	26
5 (+/- Stative)	65[10]	35

[10] Since *Joseph and Aseneth* does not use Subsystem 5, I omitted it from this calculation, rather than averaging in a zero.

The Shape of Thematization in Non-Referential Narrative

The Shape of Thematization in the Framework of Non-Referential Narrative

The normalized frequency of +Process in Tobit 4 is 33 selections/100 instances of the system. This is a little more than two standard deviations below the mean. As it turns out, however, this deviation is due simply to the predominance of projected discourse in Tobit 4, which has only three framework clauses. This observation testifies to the problems with the procedure for generating normalized frequencies discussed in Chapter 8 under "Lessons Learned."

The normalized frequency of +Circumstance in Jdt 5 is 83 selections/100 instances of the system. This falls more than three standard deviations above the mean. This points towards this chapter markedly shying away from placing the categories typically realized by nouns, which would fall under -Circumstance, in favor of thematic adjuncts. In terms of seeking a functional explanation for this striking deviation in frequencies, it is worth noting that one of the key tasks performed by thematic noun groups is shifting the scene, "perhaps signalling that a new person or event is the center of focus."[11] The one thematic subject in Judith 6 is occurs just after the slaves of Holofernes have brought Achior to the outskirts of Bethulia, at which point πᾶς ἀνὴρ σφενδονήτης διεκράτησαν τὴν ἀνάβασιν αὐτῶν (Jdt 6:12), "every slinger held back their ascent." As a bit of background, Achior was an Ammonite who had advised Holofernes against attacking the Israelites because their God would protect them (Jdt 5:5–21), prompting backlash from the other local peoples assembled there and, ultimately, Holofernes himself who could not brook a challenge to the supremacy of

[11] Porter, *Idioms*, 296.

Nebuchadnezzar (Jdt 5:23—6:4). Holofernes decides to punish Achior by sending him to the Israelites so that he will see the inability of their God to protect them firsthand and perish with them (Jdt 6:5–9). This is how the slaves of Holofernes come to be on the outside of Bethulia. When they get there, however, they hardly stand out as the unstoppable military force Holofernes has made them out to be: the Israelites successfully keep them away from the city with sling attacks, forcing them to take cover and leave Achior there (Jdt 6:12–13). The only other case in Jdt 6 where an expressed Subject precedes its verb is at the beginning of verse fourteen: καταβάντες δὲ οἱ υἱοὶ Ισραηλ ἐκ τῆς πόλεως αὐτῶν ἐπέστησαν αὐτῷ, “and when they went down from their city, the sons of Israel stood over him [i.e. Achior—NB].” Technically, this second case does not qualify as a thematic Subject because the expressed Subject for the main clause is expressed discontinuously within an embedded clause functioning as and Adjunct, which receives the thematic designation. Nevertheless, in both cases, the expressed Subject is shifting the focus towards the Israelites.

The narrative framework of Judith contains a total of twelve thematic Complements, eleven instances of +Affected and one instance of +Stative. Five of these twelve occur in two clusters. The first of these clusters, consisting of two clauses within Jdt 1:3, stresses the size of the towers Arphaxad built around his capital. The second cluster, consisting of three clauses in Jdt 2:27, stresses the extent of the destruction Holofernes and his army caused around Damascus, underscoring the surprising nature of Israel’s ultimate victory.

The Shape of Thematization of Discourse Within Non-Referential Narrative

The projected discourse portion of Tobit 9 contains no thematic Adjuncts. The resultant normalized frequency of 0 selections/100 instances of the system falls more than two standard deviations below the mean normalized frequency for +Circumstance in projected discourse within Tobit. This potentially suggests a shift towards the categories typically realized by nouns, which select -Circumstance. Although there are only two selections from subsystem 2 (+/- Circumstance) in the chapter, which would normally cause me to question the usefulness of extrapolating too much from the deviation in normalized frequency, in this case both selections of -Circumstance put the focus on participants who are crucial within the plot of the book as a whole and this specific chapter. In the first case, Tobias asks his companion to retrieve the silver from Gabael "and bring him to the wedding" (καὶ αὐτὸν ἄγε εἰς τὸν γάμον). Retrieving the silver Tobit had left on deposit with Gabael is the inciting incident for the whole book, and Gabael plays a critical part in the wedding festivities themselves by blessing the couple (Tob 9:6).[12] The second participant to receive the spotlight by being placed in thematic position is the titular character himself: the reason Tobias gives Raphael, in his guise as Azarias, for sending him on to Gabael alone is that "my father counts the days" (ὁ πατήρ μου ἀριθμεῖ τὰς ἡμέρας), so Tobias does not want to tarry any more than is necessary (Tob 9:4). This contrasts with the preceding clause (Tob 9:3), where the emphasis lies on Raguel's oath, not Raguel himself (διότι ὀμώμοκεν Ραγουηλ μὴ ἐξελθεῖν με).

[12] Rahlfs adopts the variant reading, "Tobias blessed his wife," but the majority position, following other textual witnesses, is that Gabael blessed Tobias and his wife (Schmidt, "Gabael," 861–62).

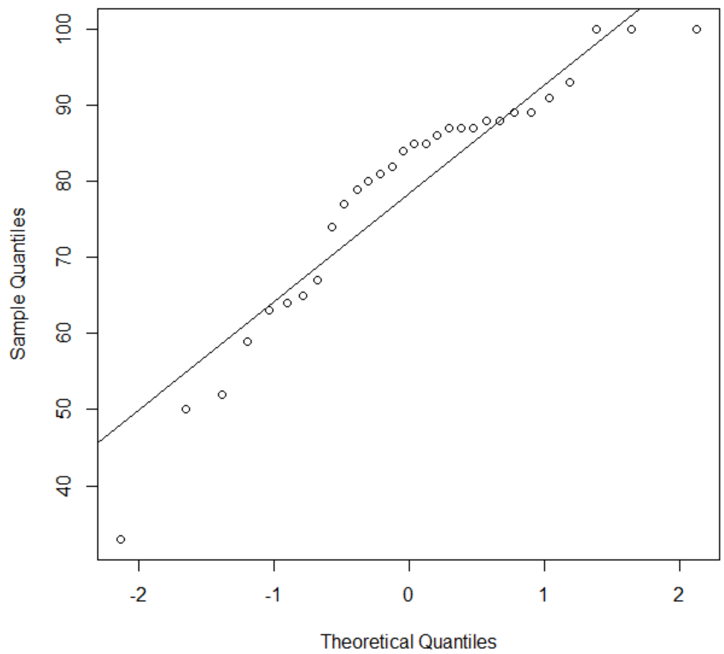
Are These Data Normal?

As with Mark and referential narrative, I need to know whether the data drawn from each subsystem conform to the normal distribution reasonably well and, thus, t-tests are an appropriate tool for analyzing them. The first subsection discuses the normality of the samples taken from the frameworks of non-referential narrative. The second subsection discusses the normality of the samples taken from the projected discourse of non-referential narrative. The third subsection summarizes the results of the previous two and their import for the further progress of the project.

As a reminder, I evaluate the data's fit to the normal distribution in four ways. First, I generate a quantile-quantile (Q-Q) plot in R and consider how closely the dots cluster around the reference line. Second, I determine the Pearson correlation coefficient (designated r) between the data sample under consideration and the normalized version R calculated as part of the Q-Q plot. Pearson's r varies between -1 (perfectly inverse relationship) to 1 (perfectly direct relationship), with 0 designating no relationship; for my purposes, I am interested in correlations stronger than r=0.92. Third, I use R to generate a histogram for the data. The histogram helps flush out false correlations where the data have two modes that look like they are close to the mean because they cancel each other out. Lastly, I report the measures of central tendency—mean, median, and mode—for the sample; these values would be precisely equal if the sample perfectly conformed to the normal distribution.

Framework Normality

The dots in the Q-Q plot (Figure 1) for the data on subsystem 1 (+/- Process) taken from the frameworks of non-referential narratives cluster quite closely around the reference



line. Note

Figure 1: Q-Q plot of Subsystem 1 (+/- Process) in framework of non-referential narratives

in particular the clustering if the extreme values are left out. Even with those values in place, the correlation coefficient is well within the bounds I set for this study (r=0.95). The histogram (Figure 2) has a single peak, although that peak is somewhat to the right of

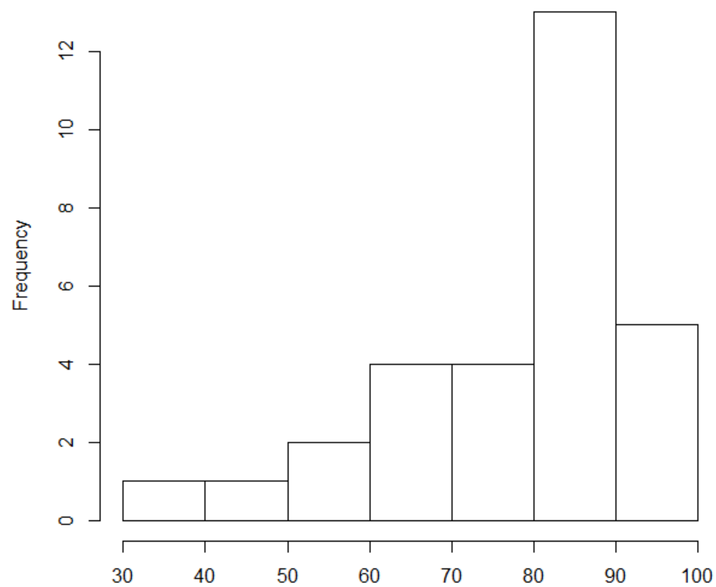


Figure 2: Histogram for subsystem 1 (+/- Process) in the frameworks of non-referential narrative

where it would be for a perfectly normal set of data. The mean is 79 selections/100 instances of the system. The median is 85 selections/100 instances of the system. The two modes, which occurred three times apiece, are 87 selections/100 instances of the system and 100 selections/100 instances of the system. Although it might seem that the presence of two modes indicates that the distribution is bimodal rather than normal, the two values for the mode are both on the same side of the mean, so the histogram does not have the valley in the middle that is characteristic of the bimodal distribution. Furthermore, 88 selections/100 instances of the system also occurred twice, so the presence of two modes could literally be a rounding error: if either occurrence of 88 selections/100 instances of the system had rounded down, 87 and 100 selections per 100 instances of the system would no longer be tied for mode status. All four of these indicators point towards using t-tests on this data. Likewise, while the magnitude of the skewness is a bit larger than one would prefer (-1.04), as the histogram indicated, the excess kurtosis is quite reasonable (0.73).

Turning to the framework's data on subsystem 2, the Q-Q plot (Figure 3) shows some clustering around the reference line. This sample initially appears to match the

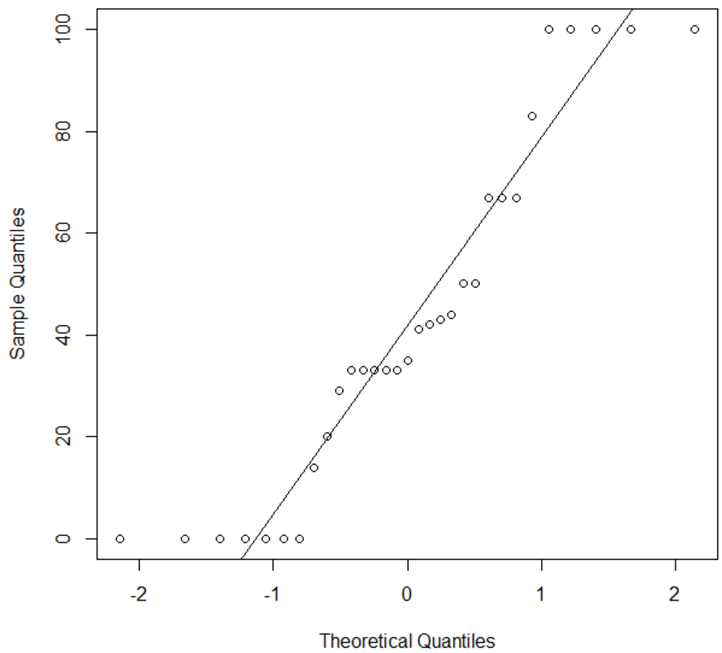


Figure 4: Q-Q plot for subsystem 2 (+/-Circumstance)

normal distribution as well as subsystem 1's (r=0.95), but the histogram (Figure 4) is

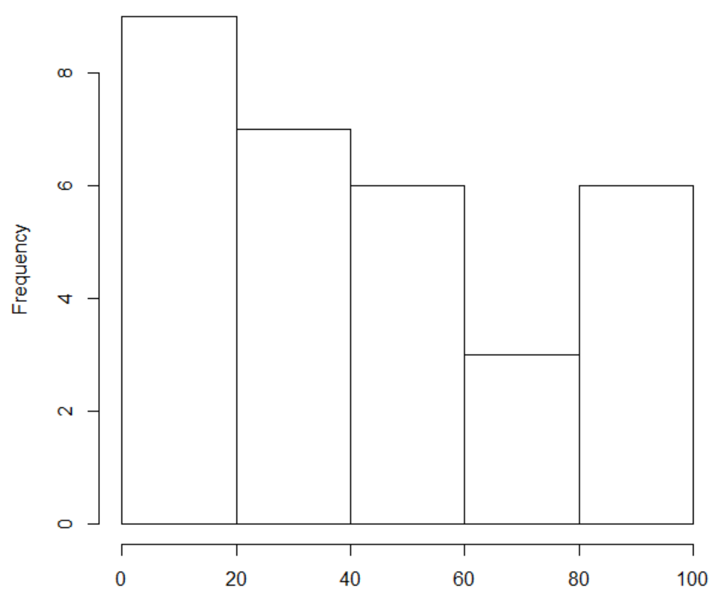


Figure 3: Histogram for subsystem 2 (+/- Circumstance) in the framework of non-referential narrative

clearly bimodal, indicating that the putative correlation is a mirage. The mean is 43 selections/100 instances of the system. The median is 35 selections/100 instances of the system. The mode is 0 selections/100 instances of the system. The skew (0.41) and kurtosis relative to that of the normal distribution (-0.88) are quite reasonable. The histogram and the spread of the measures of central tendency indicate that this sample is not normally distributed and, thus, I will not perform t-tests on it.

The Q-Q plot of the data on subsystem 3 taken from the framework of non-

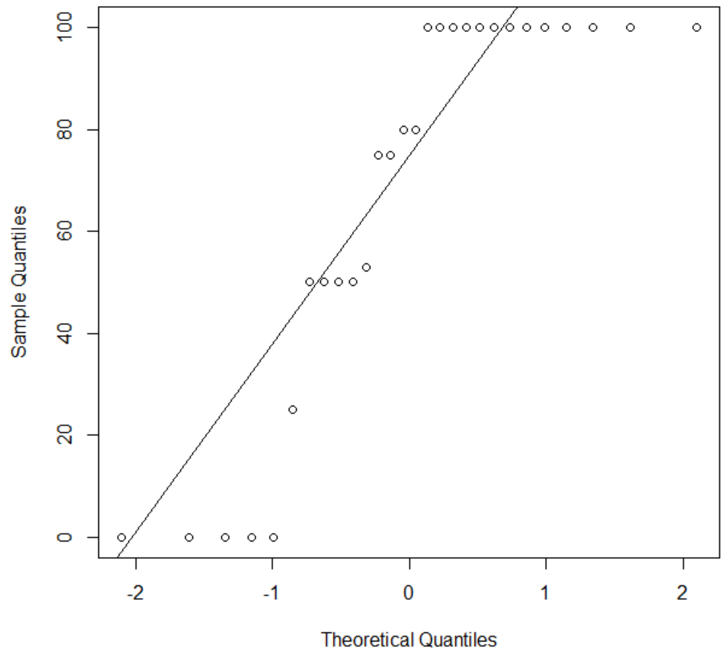


Figure 5:Q-Q plot for subsystem 3 (+/-Subject) in the frameworks of non-referential narrative

referential narratives (Figure 5) show little clustering. The correlation coefficient reinforces this impression (r=0.89). The histogram (Figure 6) has one peak, but it is located at the right extreme of the distribution rather than its middle. The mean is 67 selections/100 instances the system. The median is 80 selections/100 instances of the system. The mode is 100 selections/100 instances of the system. The skewness is 0.18, and the excess kurtosis is 0.64.

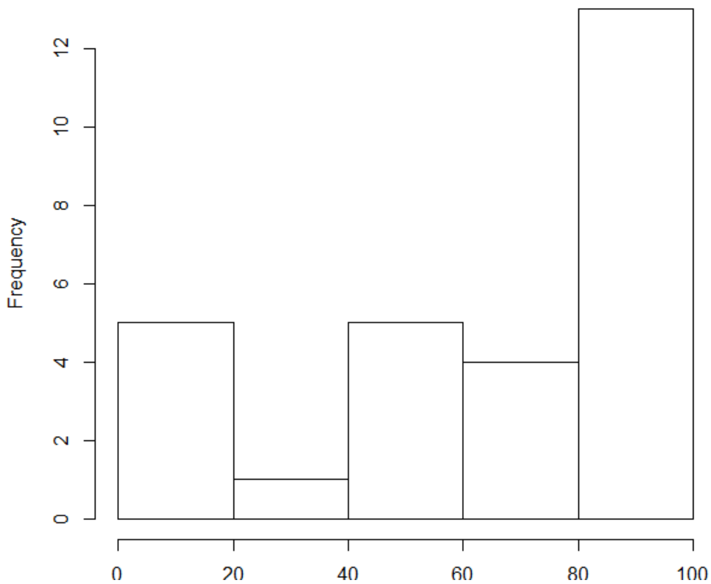


Figure 6:Histogram of subsystem 3 (+/-Subject) in the framework of referential narrative

The Q-Q plot for the data on subsystem 4 in the framework of non-referential

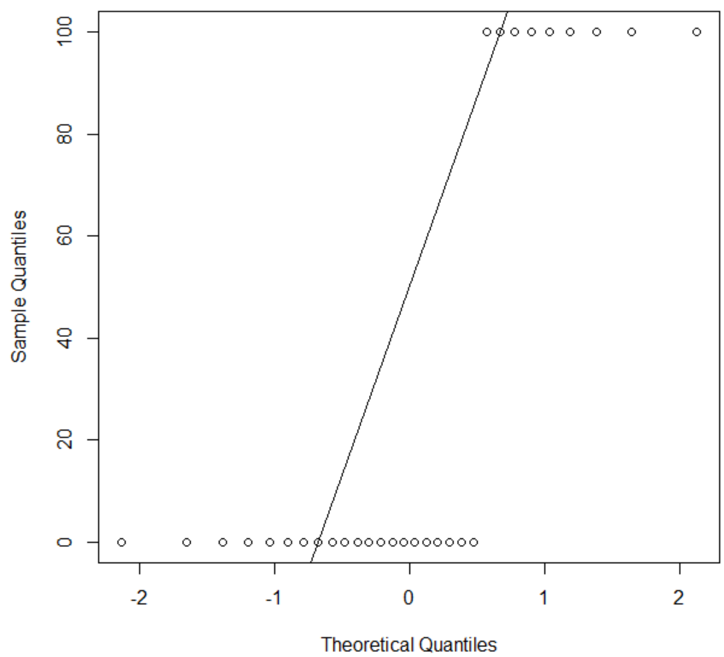


Figure 7: Q-Q plot of subsystem 4 (+/-affected) in the frameworks of non-referential narrative

narratives (Figure 7) shows how utterly non-normally distributed this sample is. The correlation coefficient reinforces this impression (r=0.77). The histogram (Figure 8) reveals a textbook example of the bimodal distribution—the two peaks are at 0 selections/100 instances of the system and 100 selections/100 instances of the system.

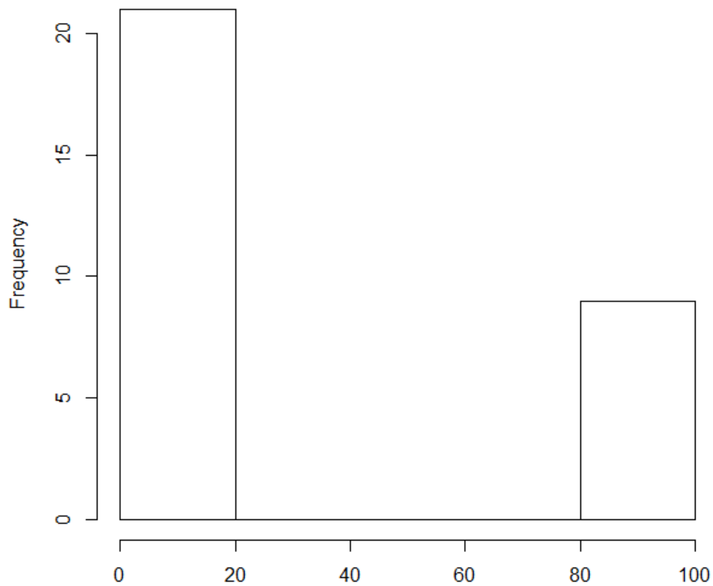


Figure 8: Histogram of subsytem 4 (+/- affected) in the framework of non-referential narrative

The mean is 30 selections/100 instances of the system. The median is 0 selections/100 instances of the system. The mode is 0 selections/100 instances of the system. The skewness of these data is 0.87, and the kurtosis relative to that of the normal distribution is -1.24. Clearly, t-tests are not admissible for this subsystem.

There is only one instance of subsystem 5 in the frameworks of my non-referential samples (Jdt.8_2). This fact makes the rest of the procedure meaningless. One cannot meaningfully speak of one data point conforming or not conforming to a theoretical distribution, so there is no point in pursuing the question further.

Discourse Normality

This subsection contains evaluations of the fit between the data taken from the discourse sections of non-referential narratives and the normal distribution. One special consideration that affected these tests is that several chapters in this sample have no dialogue. As a result, I pruned my sample to those chapters that had at least one clause of projected discourse.

Whereas the raw data for the data on subsystem 1 from projected discourse of non-referential narratives had only a borderline correlation with the normal distribution before the pruning (r=0.92), it was quite normal afterwards (r=0.96). The Q-Q plot of the pruned data (Figure 9) shows that the data cluster around the reference line quite well.

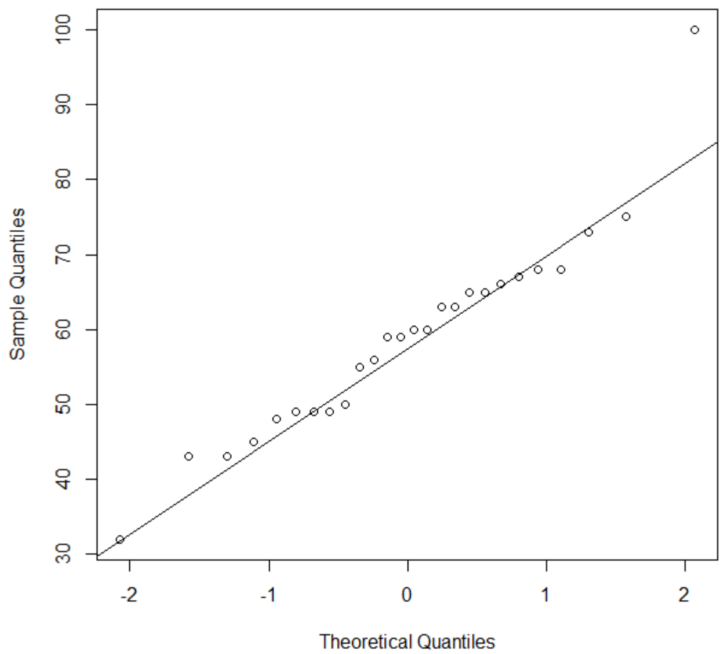


Figure 9: Q-Q plot for subsystem 1 (+/-Process) in projected discourse of non-referential narratives

The histogram of the data (Figure 10) shows that the distribution is slightly bimodal, but the peaks are close together near the center of the diagram, so this does not seem fatal to the use of t-tests here. They do, however, contribute to the large value for excess kurtosis

(1.84). The skew is quite reasonable (0.76).

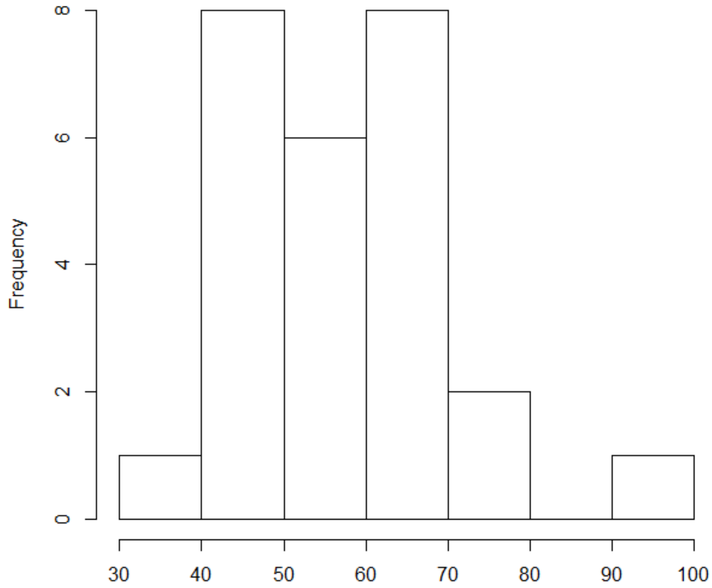


Figure 10: Histogram of subsystem 1 (+/- process) in projected discourse of non-referential narrative

The mean is 59 selections/100 instances of the system. The median is 60 selections/100 instances of the system. The mode is 49 selections/100 instances of the system.

Pruning the chapters without dialogue helped with subsystem 2 as well. The raw data had a reasonable correlation coefficient (r=0.96) but a bimodal distribution. The pruned data had an even higher quantitative correlation (r=0.98). The Q-Q plot (Figure

11) shows the majority of the data points clustering very close to the reference line.

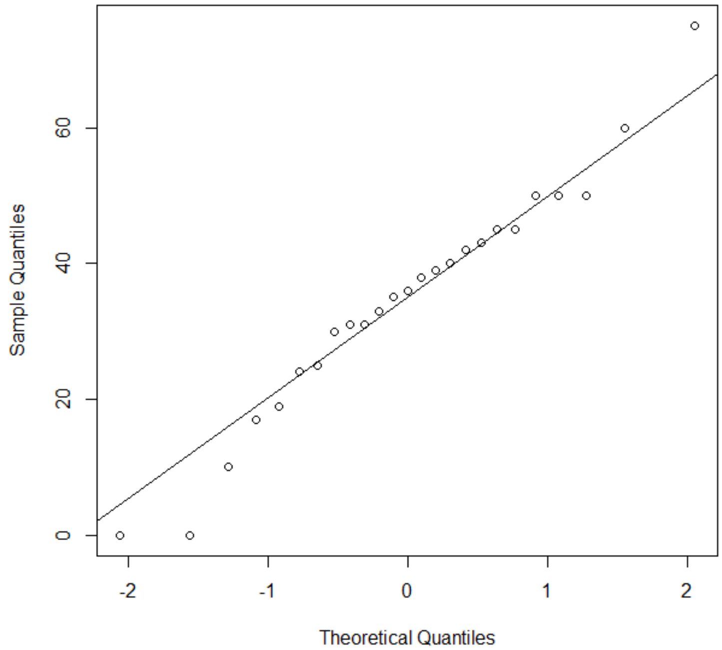


Figure 11: Q-Q plot for subsystem 2 (+/-circumstance) in projected discourse of non-referential narratives

The histogram (Figure 12) is not perfectly bell-shaped, falling off lots quicker to the left than it does to the right, but there is one peak roughly in the middle of the data.

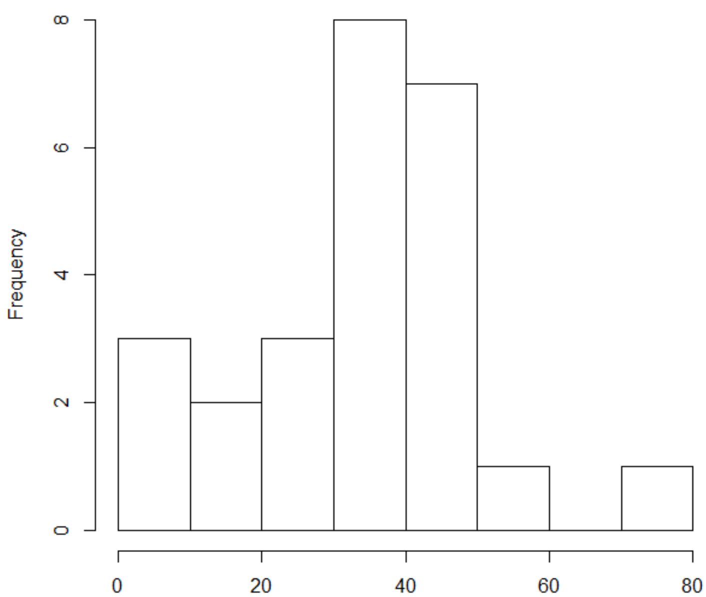


Figure 12: Histogram of subsystem 2 (+/-circumstance in projected discourse of non-referential narratives

As the histogram would indicate, the data are a bit negatively skewed (-0.12). The excess kurtosis is 0.23, indicating that the data are slightly more peaked than the normal distribution. The mean is 35 selections/100 instances of the system. The median is 36 selections/100 instances of the system. The mode is 50 selections/100 instances of the system. Although the mode differs substantially from the other measures of central tendency, the other indicators point to the applicability of t-tests, so I will perform them.

The Q-Q plot of the pruned data from non-referential narratives on subsystem 3 (Figure 13) clusters fairly well around the reference line.

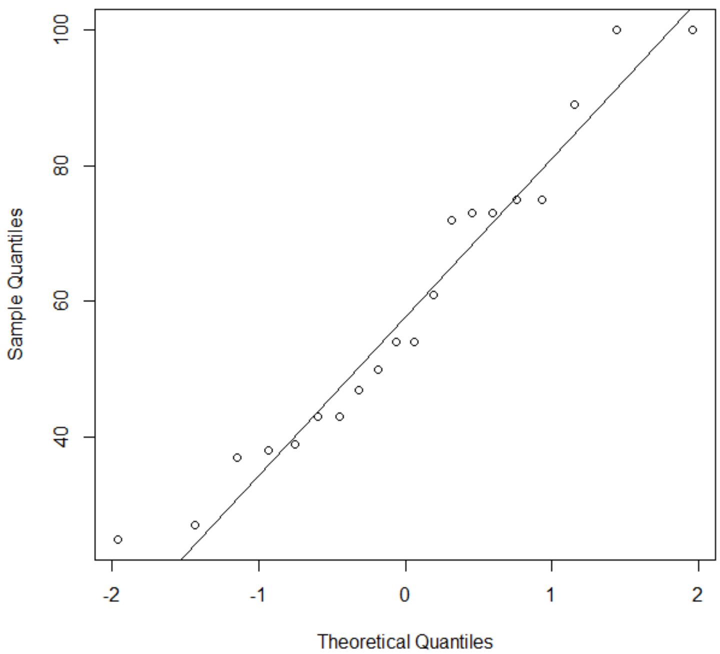


Figure 13: Q-Q plot for subsystem 3 in projected discourse of non-referential narrative

The quantitative measure of correlation is also quite high (r=0.98). The histogram (Figure 14) is borderline, coming close to being bimodal, but the higher peak is close to

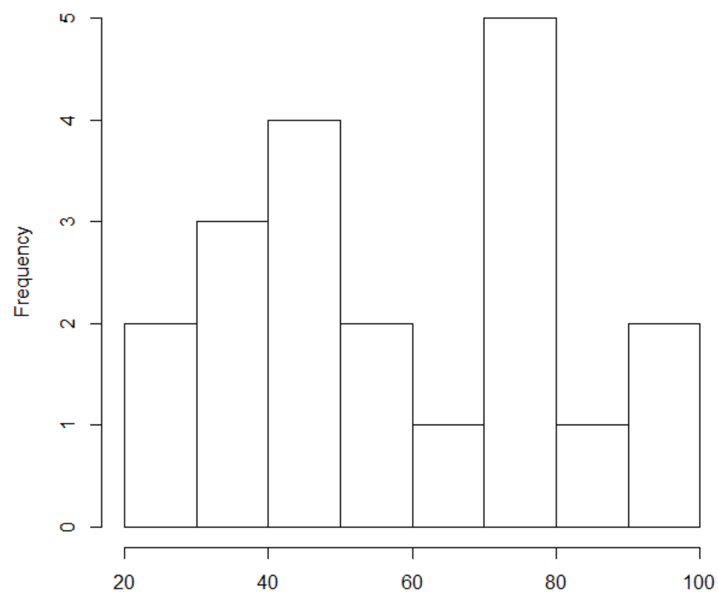


Figure 14:Histogram of subsystem 3 (+/-subject) in the projected discourse of non-referential narrative

the middle of the data. The mean is 59 selections/100 instances of the system. The median is 54 selections/100 instances of the system. Several values occur twice, tying for the status of mode, namely 43 selections/100 instances of the system, 47 selections/100 instances of the system, 54 selections/100 instances of the system, 73 selections/100 instances of the system, 75 selections/100 instances of the system, 100 selections/100 instances of the system. The data are positively skewed somewhat (0.36) and fatter than the normal distribution (the kurtosis is 0.87 less than that of the normal distribution).

The admissibility of a t-test here is open to question. The Q-Q plot and the correlation coefficient indicate that the data are normally distributed, but the histogram calls this into question. The measures of central tendency are split as well: the mean and median are quite close (54 and 59 selections/100 instances of the system, respectively), but a number of values are tied for the mode. Significantly, however, one of those values is 54 selections/100 instances of the system, the value of the mean. This means that at least part of the weight associated with the mode lands in favor of a normally-distributed set of data, leading me to include a t-test here—particularly since so many of my other tests are clearly unperformable.

The Q-Q plot of the pruned data on subsystem 4 from non-referential narratives (Figure 15) shows that the data between the extremes fit the normal distribution; the problem is that most of the data are at the extremes. The quantitative measure of the fit

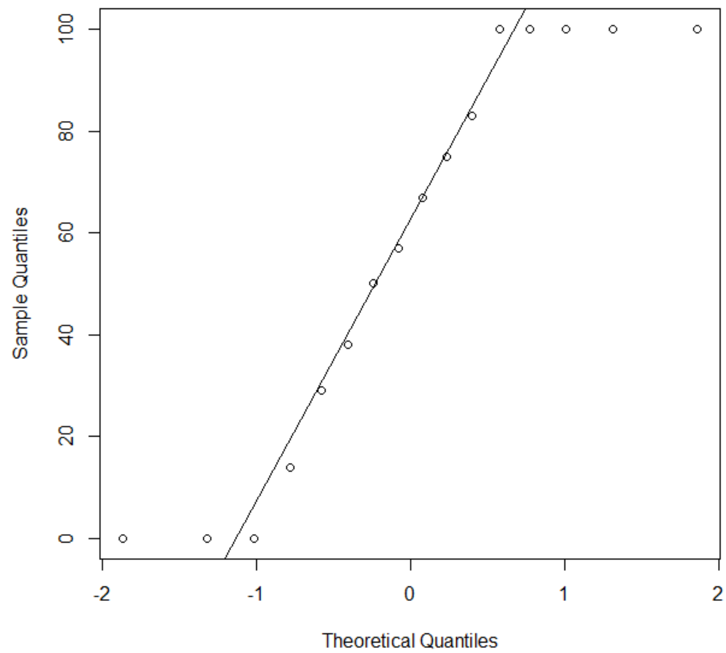


Figure 15: Q-Q plot of subsystem 4 in the projected discourse of non-referential discourse

with the normal distribution is reasonable (r=0.94). This seems to be a false positive, however, because the histogram (Figure 16) is clearly bimodal. The data are somewhat negatively skewed (-0.28), since the right peak of the histogram is higher than the left. These data are fatter than the normal distribution, for their kurtosis relative to it is -1.40. The mean is 57 selections/100 instances of the system. The median is 62 selections/100 instances of the system. The mode is 100 selections/100 instances of the system.

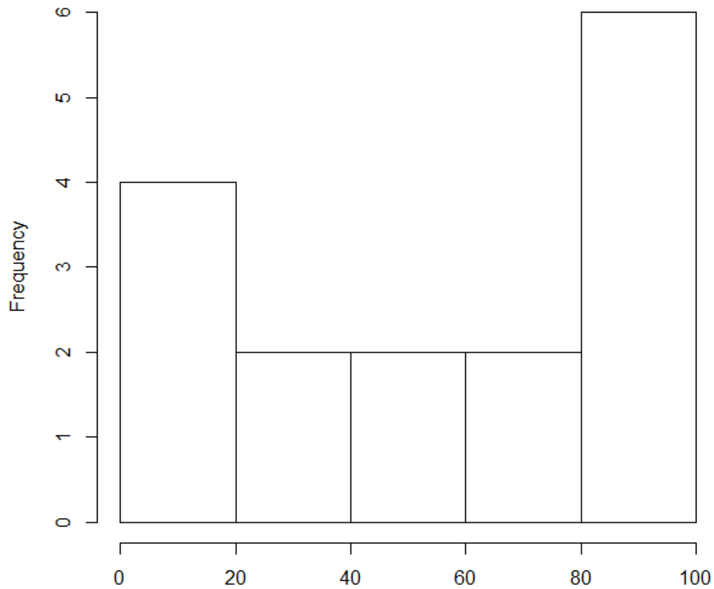


Figure 16: Histogram of subsystem 4 (+/-affected) in the projected discourse of non-referential narratives

The data for subsystem 5 within the projected discourse of non-referential narratives exist even more at extreme values than do the data from subsystem 4 (r=0.78). Here, only four values were *not* either 0 selections/100 instances of the system or 100 selections/100 instances of the system. For the sake of completeness, Figures 17 and 18 below show the Q-Q plot and histogram. The skewness of the data is quite large (1.24); the kurtosis is slightly less than that of the normal distribution (-0.19). The mean of the data is 23 selections/100 instances of the system. The median is 0 selections/100 instances of the system. The mode is 0 selections/100 instances of the system. These data are of a bimodal distribution, so t-tests are inadmissible.

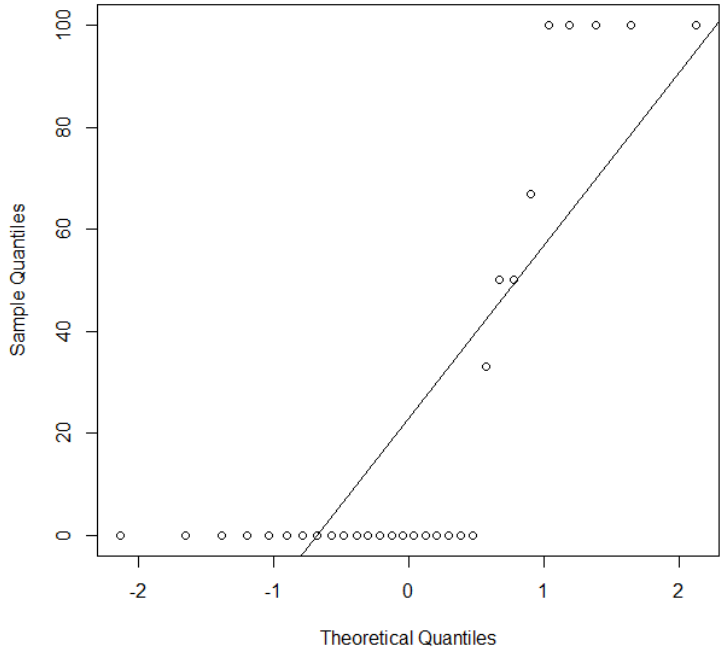


Figure 17: Q-Q plot for subsystem 5 in projected discourse of non-referential narrative

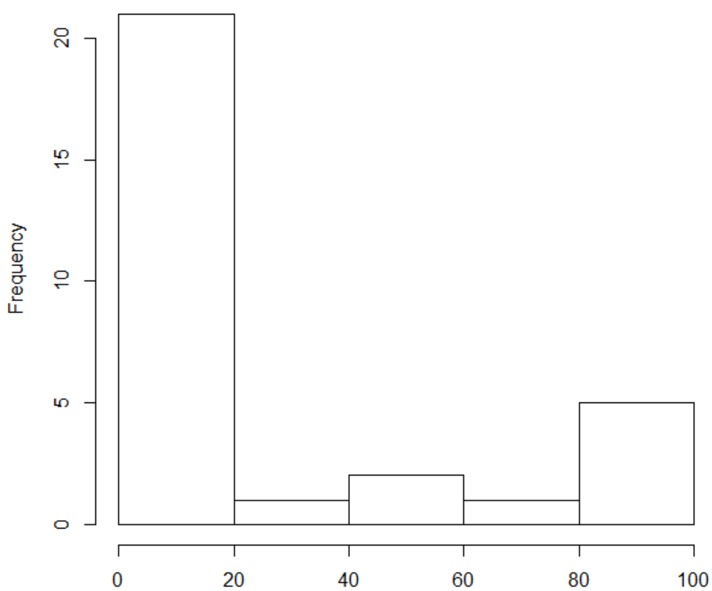


Figure 18: Histogram of subsystem 5 in projected discourse of non-referential narrative

Summary

In summary, the above analysis supports the use of t-tests for four out of ten sets of data. The distribution of the data from subsystem 1 supports a t-test for both the framework and discourse. The other two subsystems with approximately normal distributions are subsystems 2 and 3 within discourse. With these results in hand, I now proceed to Chapter 7 in which I will compare the results from Mark to both these results and those of Chapter 5.

CHAPTER 7: COMPARING MAPS

This chapter reports the results of the statistical comparisons I carried out on the data reported in chapters 4 through 6. It consists of four major parts. This introductory section previews the structure of the remainder of the chapter before reviewing some statistical concepts and procedures I introduced in Chapters 2 and 3. Section 7.1 covers comparisons between Mark and the referential portion of the corpus. Section 7.2 handles comparisons between Mark and the non-referential corpus.

Like the comparison of Mark's narrative framework and projected discourse in section 4.4., the statistical analyses reported here are hypothesis tests. As a reminder, statistical tests cannot definitively prove anything; they only prove how likely something is. This is why hypothesis tests actually use the null hypothesis (often written H_0), which is the converse of the research hypothesis: if the researcher frames his or her hypotheses in proper logically complementary fashion, then proving that the likelihood of A is less than 5 percent is the same as proving the likelihood of ~A to be over 95 percent. I will return to the wording of—and rationale for—both null and research hypotheses below as I report the results of each test. I purposefully chose the example of 5 percent and 95 percent: the significance level, i.e. the value less than which my test result must be for me to reject the null hypothesis, for this study is 5 percent; thus, if the results justify rejecting the null hypothesis, I can legitimately infer that the research hypothesis is likely to be 95 percent true. Once the hypotheses are in place, the next step is to perform the tests themselves, generating a χ^2 value and, if applicable, a t-value. Then, I must compare the value to the table of critical values for the test statistic in question. If the computed value exceeds the printed critical value, then I can legitimately reject the null hypothesis in favor of the research hypothesis. Following standard practice in quantitative research, the results below report exact values of the test statistic, the degrees of freedom involved, and the p-value associated with this result.[1]

Mark and Referential Narrative

This section covers the comparisons between Mark and the comparative referential narratives. Each subsystem potentially offers four tests—two χ^2 analyses for homogeneity and two t-tests for equal sample means. In the case of both tests, one test compares the frameworks of Mark and referential narrative and one compares projected discourse in Mark to projected discourse in the referential narratives. Five subsystems and four tests per subsystems means twenty potential tests. Unfortunately, some of these tests proved unperformable; in fact, all of the tests for Subsystem 5 were unperformable. As a result, only subsystems 1 through 4 receive a subsection below. Each subsection discusses which tests I was able to perform, why I could not perform the others, the process of each performable test, and the results generated. A final subsection then ties the results of this section together, preparatory to section 7.3, in which I compare the results of these comparisons between Mark and referential narrative and the results of section 7.2, the comparisons between Mark and non-referential narratives.

Comparing Mark and Referential Narrative in Subsystem 1 (+/-Process)

The data for this subsystem meet the assumptions for all four tests. In each case, I am interested in discovering differences between Mark and referential narratives. Therefore, the null hypothesis is that Mark's choices of ideational theme match those of referential

[1] Larson-Hall and Plonsky, "What Gets Reported," 128.

narrative. The alternative hypothesis is that they do not match. As a reminder, the chosen significance level is 5 percent (α=0.05).

The first test is χ^2 analysis of the framework. This is a χ^2 test of homogeneity, since I am interested in determining whether a difference exists between two categorical variables, namely “Mark” and “referential narrative.” If the relative proportions, i.e. the grammatical probabilities, with which Mark and referential narratives choose +/- Process are the same, then the proportion that emerges from combining these sets of data should serve as a reasonable estimate of that proportion/grammatical probability. Table 4.2 shows that Mark’s framework selects +Process 409 times and -Process 337 times for a total of 746 choices. Combining the data in Tables 5.5 and 5.6 indicates that the frameworks of referential narratives select +Process 442 times and -Process 211 times, for a total of 653 choices. This means the overall proportion for subsystem 1, considering Mark and referential narrative together, is 0.61/0.39 for +/- Process respectively. Thus, if the null hypothesis is true, 0.61/0.39 should reasonably approximate the overall grammatical probability. The expected frequencies that R computed on this assumption were 454 for +Process in Mark’s framework, 292 for -Process in Mark’s framework, 397 for +Process in the framework of referential narratives, and 256 for -Process in the framework of referential narratives. According to R, the computed value of the test statistic is 23.64 with 1 degree of freedom. Consulting a standard χ^2 table, the critical value for 1 degree of freedom and a significance level of 5 percent is 3.84.[2] Clearly, this test provides overwhelming evidence for rejecting the null hypothesis and concluding that a real difference exists between the thematization choices of the frameworks of Mark

[2] Woods et al., *Statistics*, 301.

and the comparative referential narratives with regard to subsystem 1. As a measure of exactly how strong the evidence of a difference is, R assigns a p-value of 1.16×10^{-6} to the results of this test. The p-value measures the likelihood of finding a result at least this extreme in a case where the null hypothesis is actually true. In this case, there is roughly one chance in a million that a case where the null hypothesis was true would produce samples this different.

The next test to perform is a χ^2 analysis of the choices in projected discourse. Proceeding along the same lines as previously, the aggregate totals are 441 selections of +Process and 597 selections of -Process, resulting in an overall proportion of 0.42/0.58 for +/- Process respectively. This proportion gives expected frequencies of 318 for +Process in Mark's projected discourse, 430 for -Process in Mark's projected discourse, 123 for +Process in the discourse of referential narratives, and 167 for -Process in the discourse of referential narratives. R computed the value of the test statistic as 0.27 with 1 degree of freedom. This is substantially less than the critical value of 3.84, meaning I have insufficient evidence to reject the null hypothesis. As a matter of fact, the p-value R gives for this test (0.60) indicates that the null hypothesis would be true 60 percent of the time.

Turning to the t-tests, the results R gives for comparing the framework do not support rejecting the null hypothesis (t=-1.21 for 15.14 degrees of freedom).[3] Samples like the framework of Mark and the framework of the comparative referential narratives could be randomly drawn from the same population nearly a quarter of the time (p=0.24). Comparing the discourse portions of Mark and referential narratives tells a similar story:

[3] Woods et al., *Statistics*, 300.

once again the computed t-value (-1.31 for 12.44 degrees of freedom) is insignificant, though a touch less insignificant than for the framework (p=0.21 vs. p=0.24 for the framework).

Only one of the four tests above, namely the χ^2 analysis for the framework, provided sufficient evidence to justify rejecting the null hypothesis. Technically, this does not prove that Mark and referential narratives match—just that I cannot prove that they do not.

Comparing Mark and Referential Narrative in Subsystem 2 (+/-Circumstance)

The data for this subsystem meet the assumptions for all four tests. In each case, I am interested in discovering differences between Mark and referential narratives. Therefore, the null hypothesis is that Mark's choices of ideational theme match those of referential narrative. The alternative hypothesis is that they do not match. As a reminder, the chosen significance level is 5 percent (α=0.05).

The first test compares Mark and the framework of referential narratives. Mark's narrative framework selects +Circumstance 185 times and -Circumstance 152 times, for a total of 337 choices. The narrative framework of referential narratives selects +Circumstance 81 times and -Circumstance 130 times. The overall proportion works out to 0.49/0.51 for +Circumstance and -Circumstance respectively. Using this overall proportion as a basis, R computes expected frequencies of 164 for +Circumstance and 173 for -Circumstance in Mark's framework and 102 for +Circumstance and 109 for -Circumstance in the frameworks of referential narrative. According to R, the computed value for the test statistic is 13.50 with 1 degree of freedom. The critical value for 1 degree of freedom is 3.84, so this result is highly significant (p=0.00023).

The next test is a χ^2 analysis of the choices in projected discourse. The projected discourse sections of Mark select +Circumstance 140 times and -Circumstance 286 times. The comparative referential narratives select +Circumstance 52 times and -Circumstance 119 times. The overall proportion is 0.32/0.68 for +Circumstance and -Circumstance respectively. According to R, the resultant expected frequencies are: 137 selections of +Circumstance in Mark, 289 selections of -Circumstance in Mark, 55 selections of +Circumstance in referential narratives, and 116 selections of -Circumstance in referential narratives. R's value for the test statistic is 0.23 for 1 degree of freedom. The critical value of a χ^2 for 1 degree of freedom is 3.84, so the computed value of the test statistic does not support rejecting the null hypothesis ($p=0.63$).

I also performed separate t-tests comparing Mark's framework to that of referential narratives. The results for the framework barely support rejecting the null hypothesis ($t=2.21$ for 24.67 degrees of freedom); there is a chance of roughly 4 percent that these samples come from the same population ($p=.04$). The results for discourse, however, offer no solid case for rejecting the null hypothesis ($t=0.18$ for 21.35 degrees of freedom). Samples like these could occur randomly over 85 percent of the time ($p=0.86$).

Both of the framework tests for this subsystem supported rejecting the null hypothesis; both the discourse tests did not. Although the significant χ^2 result was highly significant (the p-value R assigned the result was less than 1 percent of the cut-off value), the significant t-test result was borderline ($p=0.04$ vs. $\alpha=0.05$). Overall, these data seem to be fairly consistent with the null hypothesis that Mark's choices of ideational theme correspond to those of referential narratives, but the case they make in that direction is not nearly as strong as the one the data from subsystem 1 make.

Comparing Mark and Referential Narrative in Subsystem 3 (+/- Subject)

The analyses of normality indicated some concerns about the normality of the data, making t-tests unreliable. The rarity of thematic Complements of all types (+Affected, +Stative, and -Stative) in the frameworks of both Mark and referential narratives means that several chapters within both Mark and the referential narratives have no instances of -Subject, resulting in a normalized frequency of 100 instances per 100 choices from the system for +Subject. This value is, in fact, the mode for both sets of framework data. The hallmark of the normal distribution is that the most common values for the data should be somewhere in the middle of the range, so cases like these where the most common value for the data is at an extreme clearly do not resemble the normal distribution. The data from referential narrative would look normal if I simply ignored these extreme values as outliers. On the other hand, treating the most common value as an outlier seems questionable from a methodological perspective even before considering that I would need to do the same to Mark, where so-called outliers would constitute a full half of the data points. Turning to discourse, the data from Mark seem reasonably close to the normal distribution, the data from referential narrative do not. On the other hand, these data meet the assumptions for the two χ^2 analysis.

The null hypothesis for the first χ^2 analysis is that an overall proportion derived from combining the framework data on subsystem 3 from Mark and referential narratives will match the proportions of the individual parts. The alternative hypothesis is that the overall proportion will differ from the individual observed proportions. Once again, the significance level is 5 percent (α=0.05). Mark selects +Subject 139 times and -Subject 13 times in his framework, for a total of 152. The frameworks of referential narratives include 102 instances of +Subject and 28 of -Subject, for a total of 130. The overall proportion derived from combining these data is 0.85 for +Subject to 0.15 for -Subject. According to R, the expected frequencies generated from this overall proportion are 130 for +Subject and 22 for -Subject for Mark's framework and 111 for +Subject and 19 for -Subject in the frameworks of referential narrative. R's computed value for the χ^2 test statistic is 8.49 with 1 degree of freedom. The critical value of χ^2 for 1 degree of freedom is 3.84, so this result justifies rejecting the null hypothesis (p=0.0036).

The discourse χ^2 analysis uses hypotheses analogous to the framework test, and the significance level matches as well (α=0.05). The discourse portions of Mark contain 188 instances of +Subject and 98 instances of -Subject, while the discourse portions of referential narratives contain 76 instances of +Subject and 43 instances of -Subject. This means 264 total instances of +Subject and 141 instances of -Subject, for an overall proportion of 0.65 for +Subject to 0.35 for -Subject. The expected frequencies emerging from this overall proportion are, according to R: 186 for +Subject in Mark, 100 for -Subject in Mark, 78 for +Subject in the discourse portions of referential narratives and 41 for -Subject in the discourse portions of referential narratives. These expected frequencies are extremely close to the observed ones, so it should be no surprise that R computed a microscopic value for the χ^2 test statistic (0.06 for 1 degree of freedom). Once again, the critical value of χ^2 for 1 degree of freedom is 3.84; this result clearly does not justify rejecting the null hypothesis. In fact, the p-value associated with this result indicates that the null hypothesis will be true roughly four out of every five times (p=0.81).

The two performable tests on the data from this subsystem point in opposite directions. The χ^2 analysis of the framework data returned a solidly significant result

(χ^2=8.49 for 1 degree of freedom, p=0.0036), while the χ^2 analysis of the discourse failed to reject the null hypothesis (χ^2=0.06 for 1 degree of freedom).

Comparing Mark and Referential Narratives in Subsystem 4 (+/-Affected)

The data from this subsystem support only one test, namely a χ^2 analysis of the discourse data. Neither the data from Mark nor referential narrative passed the tests of normality, which eliminates the use of t-tests. Additionally, Mark contains only 2 instances of the choice -affected, violating the condition of χ^2 analysis that all cells must contain a frequency of at least 5. This leaves only the χ^2 analysis of discourse data.

The null hypothesis for this test is that the overall proportion derived from combining the data from Mark and referential narratives will match the proportions derived independently. The alternate hypothesis is that the proportions will differ. The discourse portions of Mark contain 70 instances of +Affected and 28 instances of -Affected. The discourse portions of referential narratives contain 32 instances of +Affected and 11 instances of -Affected. The totals of 102 instances of +Affected and 39 instances of -Affected result in an overall proportion of 0.73 for +Affected to 0.27 for -Affected. The expected frequencies computed on the basis of the null hypothesis are: 71 for +Affected in Mark's discourse, 27 for -Affected in Mark's discourse, 31 for +Affected in the discourse of referential narratives, and 12 for -Affected in the discourse of referential narratives. The observed and expected frequencies match almost exactly in this case, generating an extremely small χ^2 value (0.026 with 1 degree of freedom). Comparing this value to the critical value of χ^2 analysis for 1 degree of freedom and 5 percent significance, which is 3.84, shows that this result does not justify rejecting the null hypothesis.

The Overall Picture of Comparing Mark and Referential Narratives

In theory, this section could have contained twenty tests—five subsystems with four tests per subsystem. Of these, nine (all four of the tests for subsystem 5, three of the four for subsystem 4, and the two t-tests for subsystem 3) were unperformable because the data I collected did not meet their assumptions. Five of the remaining eleven offered sufficient evidence to overturn the null hypothesis, although one of those, the framework t-test for subsystem 2, was borderline. The other six did not offer evidence substantial enough to overturn the null hypothesis.

Mark and Non-referential Narratives

This section covers the comparisons between Mark and the comparative non-referential narratives. Each subsystem potentially offers four tests—two χ^2 analyses for homogeneity and two t-tests for equal sample means. In the case of both statistical tools, one test compares the frameworks of Mark and referential narrative and one compares projected discourse in Mark to projected discourse in the referential narratives. Five subsystems and four tests per subsystems means twenty potential tests. Unfortunately, I could not perform most of the t-tests because only four of the ten groups of data passed the tests for normality described in section 6.6 and three of the ten groups, namely the framework data from subsystem 4 and both groups of data for subsystem 5, occur too rarely to support reliable χ^2 analyses. None of the performable t-tests happen to be for categories that do not support reliable χ^2 analyses, meaning that no tests are performable for subsystem 5. As a result, only subsystems 1 through 4 receive a subsection below. Each subsection discusses which tests I was able to perform, why I could not perform the others, the process of each performable test, and the results generated. A final subsection then ties the results of this section together.

Comparing Mark and Non-referential Narratives in Subsystem 1 (+/- Process)

This is the only subsystem that supports all four tests. I am interested in finding differences between Mark and non-referential narratives, so the null hypothesis in each case is that the grammatical probabilities of Mark's ideational theme choices matches the choices of non-referential narrative. In statistical terms, for χ^2 analyses, the relative proportions, i.e. grammatical probabilities, derived from combining the data from Mark and non-referential narratives should match the proportions observed individually.

The first test is a χ^2 analysis comparing the framework of Mark and non-referential narratives. Mark selects +Process 409 times and -Process 337 times, for a total of 746 choices. Non-referential narratives choose +Process 567 times and -Process 159 times, for a total of 726 choices. This means the overall proportion is 0.66/0.34 for +/-Process respectively. If the null hypothesis is true, this proportion should reasonably approximate the individual probabilities. The expected frequencies R generated on the basis of this assumption were: 495 for +Process in Mark's framework, 251 for -Process in Mark's framework, 481 for +Process in the framework of non-referential narratives and 245 for -Process in the framework of non-referential narratives. According to R, the computed value of the χ^2 test statistic is 88.16 for 1 degree of freedom. The critical value of χ^2 for 1 degree of freedom is 3.84. This value is extremely significant ($p < 2.2 \times 10^{-16}$).

The second test is a χ^2 analysis comparing the discourse of Mark and non-referential narrative. The discourse sections of Mark select +Process 322 times and -Process 426 times. The discourse sections of non-referential narrative select +Process 428 times and -Process 328 tomes. This works out to an overall proportion of 0.49 for +Process to 0.51 for -Process. The expected frequencies calculated on the basis of this proportion are: 373 for +Process in Mark's discourse, 374 for -Process in Mark's discourse, 377 for +Process in nonreferential discourse, 379 for -Process in nonreferential discourse. The computed value for the χ^2 test statistic is 26.89 with 1 degree of freedom. The critical value of χ^2 with 1 degree of freedom and $\alpha=0.05$ is 3.84. Thus, the result is quite significant ($p=2.16 \times 10^{-7}$).

Based on the earlier tests of normality, I ran t-tests on this subsystem. The t-test on the framework offered very solid evidence for rejecting the null hypothesis in favor of concluding that a real difference exists between Mark's instantiation of subsystem 1 in the narrative framework and that of non-referential narrative (t=-6.26 for 42.00 degrees of freedom). Random draws from the same population would produce samples like these two about one time in 10,000,000 ($p=1.7\text{x}10^{-7}$). The discourse t-test, however, did not quite warrant jettisoning the null hypothesis (t=-1.63 for 41.59 degrees of freedom).

I was able to perform all four tests on this subsystem. Three of the four offered sufficient evidence to overturn the null hypothesis. Even the results of the one test that did not justify rejecting the null hypothesis were close to doing so ($p=0.11$ vs. $\alpha=0.05$). Thus, the data from subsystem 1 indicate a real difference between Mark and the comparative corpus of non-referential narrative.

Mark and Non-Referential Narratives in Subsystem 2 (+/- Circumstance)

I was able to perform three of the four tests for this subsystem. The exception was the t-test on the narrative framework, for this group of data did not pass the normality tests. The null hypothesis for the remaining t-test is that the population means of Mark and non-referential narrative with respect to the use of Circumstances are equal; the alternative hypothesis is that they differ. The null hypotheses for the χ^2 analyses are that the overall proportion derived from combining the data on subsystem 2 will match the proportions observed individually; the alternative hypotheses are that they will not.

The first test to perform is the χ^2 analysis for the narrative framework. Mark's framework selects +Circumstance 185 times and -Circumstance 152 times. The framework of nonreferential narratives select +Circumstance 81 times and -Circumstance 130 times. Combining these samples results in an overall proportion of 0.49 for +Circumstance to 0.51 for -Circumstance. R computed expected frequencies from this overall proportion: 164 +Circumstance in Mark's framework, 173 -Circumstance in Mark's framework, 102 for +Circumstance in the frameworks of non-referential narratives, and 109 for -Circumstance in the frameworks of non-referential narratives. R computed a value of 13.50 with 1 degree of freedom for the χ^2 test statistic. The critical value of χ^2 for 1 degree of freedom and a significance level of 5 percent is 3.84, so a value of 13.50 certainly justifies rejecting the null hypothesis ($p=0.002$).

The next test is a χ^2 analysis of the discourse portions of Mark and nonreferential narrative. In projected discourse, Mark selected +Circumstance 140 times and -Circumstance 286 times. The projected discourse portion from the samples of non-referential narrative selected +Circumstance 116 times and -Circumstance 212 times. This works out to an overall proportion of 0.33 for +Circumstance to 0.67 for -Circumstance. The expected frequencies R computed on the basis of this proportion are: 145 +Circumstance in Mark's discourse, 281 -Circumstance in Mark's discourse, 111 +Circumstance in discourse sections of the non-referential samples, and 217 -

Circumstance in discourse sections of the non-referential samples. The computed value of the test statistic here is 0.41 with 1 degree of freedom, which is clearly less than 3.84, which is the critical value for 1 degree of freedom and a significance level of 5 percent. Thus, the data do not support rejecting the null hypothesis. In fact, the associated p-value is just over ten times the chosen cut-off point ($p=0.52$).

The discourse data on this subsystem allow for a meaningful t-test. I performed this test in R using the version of the non-referential data that excluded chapters were there was no projected discourse. The resultant value for t was miniscule ($t=-0.24$ for 31.22 degrees of freedom). Two samples with this level of difference could come from the same population over eighty percent of the time ($p=0.81$). Clearly, this offers no support for rejecting the null hypothesis.

Two of the three performable tests on these data failed to reject the null hypothesis. The one test that did, however, did so resoundingly. Overall, the data would seem to point towards Mark and the authors of non-referential narratives thematizing their narratives similarly.

Comparing Mark with Non-Referential Narratives in Subsystem 3 (+/-Subject)

The data in this category allow for three of four tests—the two χ^2 analyses and a t-test on the discourse portion. The null hypothesis for each of the χ^2 analyses is that the overall proportion derived from combining the data from Mark and non-referential narratives will match the proportions identified individually; the alternate hypothesis is that they will not match. The null hypothesis for the t-test is that the mean frequency of the choice +Subject in Mark’s population equals the mean frequency of the choice +Subject in the population of non-referential narratives; the alternate hypothesis is that the means differ.

The first test is the χ^2 analysis comparing the frameworks. Mark's framework selects +Subject 139 times and -Subject 13 times. The frameworks of non-referential narrative select +Subject 72 times and -Subject 21 times. The resulting overall proportion is 0.87 for +Subject to 0.13 for -Subject. On the basis of this proportion, R computed the following expected frequencies: 139 for +Subject in Mark's framework, 21 for -Subject in Mark's framework, 80 for +Subject in frameworks of the samples of non-referential narratives, and 13 for -Subject in frameworks of the samples of non-referential narratives. R returned 8.36 for the value of the χ^2 test statistic, with 1 degree of freedom. Since the relevant critical value of χ^2 is 3.84, this result clearly supports rejecting the null hypothesis ($p=0.0038$ vs. $\alpha=0.05$).

The next test is a χ^2 analysis comparing the discourse portions of Mark and non-referential narratives. The discourse portions of Mark contain 188 instances of +Subject and 98 instances of -Subject. The discourse portions of non-referential narratives 128 instances of +Subject and 84 instances of -Subject. The resulting overall proportion is 0.63 +Subject to 0.37 -Subject. This proportion allowed R to compute the following as expected frequencies: 181 for +Subject in Mark's discourse, 105 for -Subject in Mark's discourse, 135 for +subject in the discourse portions of samples of non-referential narrative, and 77 for -Subject in the discourse portions of samples of non-referential narrative. The computed value for the χ^2 test statistic is 1.28 with 1 degree of freedom. This value is substantially less than the critical value of χ^2 for 1 degree of freedom and a significance level of 5 percent. Thus, the null hypothesis cannot be rejected. The probability value associated with the null hypothesis is roughly five times the chosen cut-off point for this project ($p=0.26$).

The discourse data also allow for a t-test. It comes nowhere near overturning the null hypothesis (t=0.60 for 33.80 degrees of freedom). In fact, roughly half the time (p=0.55) random chance would be sufficient to create the differences observed between the data drawn from Mark and the data drawn from non-referential narrative.

The three tests I was able to perform paint a familiar picture. The χ^2 analysis of the frameworks indicated substantial differences between Mark and non-referential narratives. On the other hand, both the χ^2 analysis and t-test of the discourse did not offer sufficient evidence to disprove that Mark and non-referential narratives were similar. Neither of these non-significant results were borderline cases (p=0.26 and p=0.55). These results seem to indicate that the choices of ideational theme in the projected discourse portions of Mark and non-referential narratives are—at the very least—not radically different.

Comparing Mark with Non-Referential Narratives in Subsystem 4 (+/-Affected)

The data allow for one test on the data from this subsystem, namely a χ^2 analysis comparing the discourse portions of Mark and non-referential narratives. The null hypothesis is that the proportion derived from combining the data taken from Mark and non-referential narratives will match the proportions observed individually; the alternative hypothesis is that the observed, individual proportions will differ significantly from the overall proportion. Mark's projected discourse contains 70 instances of +Affected and 28 instances of -Affected. The discourse sections of the samples of non-referential narratives contain 50 instances of +Affected and 35 instances of -Affected. This works out to an overall proportion of 0.66/0.34 in favor of +Affected. Using this proportion, R generated these expected frequencies: 64 for +Affected in Mark's discourse, 34 for -Affected in Mark's discourse, 56 for +Affected in non-referential narratives' discourse, and 29 for -Affected in non-referential narratives' discourse. The computed value for the χ^2 test statistic is 2.67 with 1 degree of freedom. This value is less than the critical value of χ^2 for 1 degree of freedom and a significance level of 5 percent, so I cannot reject the null hypothesis. I will note, however, that this result is closer to being significant than many of the other tests that failed to justify rejecting the null hypothesis ($p=0.10$).

The Overall Picture of Comparing Mark with Non-Referential Narrative

I was able to perform just over half of the tests that were theoretically possible (11 out of 20). Of the eleven performable tests, five supported rejection of the null hypothesis and six did not. Four of these significant results were from tests on the framework; the other was the result of the χ^2 analysis comparing the data from subsystem 1 in Mark's discourse sections to the discourse sections of non-referential narrative. This result is particularly interesting because it is the only one of the twelve performable tests on the discourse portions whose results justified rejecting the null hypothesis. This unusual result points towards a real difference between Mark and non-referential narrative in subsystem 1. In the cases of subsystems 2 through 4, however, the number—and, more particularly, the distribution—of unperformable tests makes comparison difficult. Each of the other subsystems had at least one unperformable test; subsystem 4 had three. The results of the tests I could perform on the data from these subsystems seemed to be more consistent with similarity between thematization in Mark and non-referential narrative than they were with the hypothesis that a radical difference existed between them.

Which Analogy is Better? Analyzing the Results

How does the morass of data contribute to placing the Gospel of Mark within the literary milieu of the first-century Greco-Roman world? The summaries of the preceding two sections illustrate the difficulties involved in answering this question. Both the referential and non-referential narrative sections produced eleven performable tests, a little over half the theoretical total of twenty. Obviously, if all the unperformable tests had been performable with results pointing in the same direction, the resulting picture would have been quite different. Fortunately, however, the non-performable tests for both categories are concentrated in the more delicate subsystems, rather than distributed randomly: subsystem 1 is unscathed, and subsystem 2 is nearly so (one unperformable test, the t-test comparing the frameworks of Mark and non-referential narrative, out of eight total between two the candidate analogies).

Table 7.1 below presents the results of the preceding sections side-by-side. Each test had one of three possible results: (1) the result justifies rejecting the null hypothesis, indicating a real difference between Mark and the component of the comparative corpus under discussion, (2) the result does not justify rejecting the null hypothesis that any observed difference between Mark and the component of the comparative corpus under discussion is purely random, or (3) a result was unattainable because the data did not meet the test’s assumptions. A cell in Table 7.1 corresponds to each of these tests. The code in each cell corresponds to the three possible results: “Y” indicates a result that justified rejecting the null hypothesis, “N” indicates a result that did not justify rejecting the null hypothesis, and “N/A” indicates an unattainable result.

Table 7.1 Summary of Hypothesis Test Results

Subsystem	Referential Narrative				Non-Referential Narrative			
	Frame X^2	Frame T-Test	Disc X^2	Disc T-Test	Frame X^2	Frame T-Test	Disc X^2	Disc T-Test
1	Y	N	N	N	Y	Y	Y	N
2	Y	Y[4]	N	N	Y	N/A	N	N
3	Y	N/A	N	N/A	Y	N/A	N	N
4	N/A	N/A	N	N/A	N/A	N/A	N[5]	N/A

At first glance, the table loosely points towards a correlation between Mark and referential narrative. I disproved the null hypothesis they were similar fewer times than I did for Mark and non-referential narrative (4 Y's for referential vs. 5 Y's for non-referential). Considering the results on a subsystem by subsystem basis paints much the same picture. Subsystem 1, the only one for which all the tests were performable, clearly points to a similarity between Mark and referential narrative. Only one of the four tests comparing Mark and referential narrative in subsystem 1 supported rejecting the null hypothesis; conversely *three* of the four tests involving Mark and non-referential narrative supported rejecting the null hypothesis, including the only discourse test in the entire project to do so. The importance of that last point cannot be overestimated—particularly in light of the corresponding test of Mark and referential narrative being one of the ones for which R returned a high probability for the null hypothesis. The other subsystems are mostly mixed bags with too many unperformable tests to present a clear case. Subsystem 4 is a partial exception in that both analogies offered one performable test, both of which were not significant, but the referential one returned a higher probability value.

[4] This significant result is borderline ($p=0.04$ vs. $\alpha=0.05$).
[5] This nonsignificant result is borderline ($p=0.10$ vs. $\alpha=0.05$).

A cut-and-dried, black-and-white approach to the results of null hypothesis significance tests is insufficient, for it fails to take into account the arbitrariness inherent in choosing a significance level: the only difference between p-values of 0.049999 and 0.050001 is that 0.05 has been chosen as a cut-off. In other words, should the significant result for the t-test comparing how the frameworks of Mark and referential narratives use subsystem 2 (+/- Circumstance), which returned a p-value just below the chosen cut-off (0.04), count as heavily in favor of a difference between Mark and referential narratives as the corresponding test for subsystem 1 (+/- Process), which retuned a p-value just over five times as large (0.21), counts against such a difference? I think not.

A logical next step would be to consider the raw p-values themselves, instead of only considering the p-value's position relative to the arbitrary cut-off. Unfortunately, this is not sufficient either, for the p-values are not all based on the same amount of information and, as such, should not receive equal weight. The distributional differences for two populations are, by definition, constant, but the degree to which a sample matches its population is not. This implies that the p-value of a given test depends on at least two factors, namely the constant difference between the population distributions being compared and the variable degree to which the samples used accurately represent their respective populations. Longer samples are more likely to represent their populations accurately as the random fluctuations that are an inescapable part of the sampling process average out, so longer samples return smaller p-values as the invariant difference between two populations becomes more apparent in the samples used because the samples become ever closer approximations of the populations in question. At least two inferences follow from recognizing how sample size affects p-values. First, this recognition further clarifies why hypothesis testing cannot be a simple, yes/no assessment of whether the p-value is lower than the significance level: p-values higher than the cut-off value do not necessarily prove that no significant difference exists between the populations; they can also indicate that one or both of the samples were too imprecise for a clear picture of the populations they represented to emerge. Second, p-values from a hypothesis test should not be compared directly, for they are almost certainly not based on precisely the same number of instances.

Thus, a methodologically rigorous analysis of the results for the hypothesis tests reported earlier in the chapter needs a tool that takes sample size out of the equation. The mathematical tools that fit the bill are called, intuitively enough, measures of effect size. Sample size affects different hypothesis tests differently, so each hypothesis test has its own measure(s) of effect size. The measures of effect size I have chosen, described more fully in Chapter 3, are Cramer's *V* for χ^2 analysis and Cohen's *d* for t-tests.

The results of these tests are reported in two tables below, one for referential narrative and one for non-referential narrative. The number within each cell below is the effect size calculation for the subsystem designated by the row and the test designated by the column. To facilitate the interpretation of these results without constant recourse to the discussion in Chapter 3, a parenthetical letter marks the bracket into which each result falls for the statistic in question: N for "negligible," S for "small," M for "medium," or L for "large." A discussion of each table follows it; the next subsection compares the tables.[6]

[6] As a reminder, Cramer's *V* and Cohen's *d* are not on the same scale. Therefore, the comparisons of these results below proceed based on the letters designating the brackets to which the numbers belong, not the numbers themselves.

Table 7.2: Effect Size Measurements for Referential Narrative

Subsystem	Cramer's *V* (Framework)	Cramer's *V* (Discourse)	Cohen's *d* (Framework)	Cohen's *d* (Discourse)
1	0.13 (S)	0.01610615 (N)	0.52 (M)	0.58 (M)
2	0.16 (S)	0.01979063 (N)	0.82 (L)	0.07 (N)
3	0.17 (S)	0.01459567 (N)	N/A	N/A
4	N/A	0.01355543 (N)	N/A	N/A

How Close Are the Referential Narratives to Mark?

The eleven performable tests clearly show that the data from referential narratives match the data from Mark quite well. Roughly half of the tests (5 out of 11) indicate a negligible difference between the two; half of the others (3 out of 6) indicate a small difference. As it happens, all three of the remaining results indicating a more substantial difference are Cohen's *d* scores on the lower side of their respective brackets. In fact, both the result for the framework of subsystem 1 (+/- Process) and subsystem 2 (+/- Circumstance) are only .02 standard deviations from the boundary for small and medium effects, respectively. The last result, the Cohen's *d* for the use of subsystem 1 (+/- Process) in projected discourse, lies .08 standard deviations outside thee small bracket. It is worth remembering in this connection that R computed Cohen's *d*, which is a parametric test, from the normalized frequencies.[7] Although the tests of normality for each of these groups of data showed that they conformed reasonably well to the normal distribution (otherwise, I would not be performing t-tests and calculating effect sizes for them with

[7] I had to use the normalized frequencies because the cohen.d function takes its input as vectors, and the normalized frequencies were in vector form. See "Lessons Learned" in Chapter 8 for an explanation of how a better procedure for data collection would have allowed me to produce a vectorized form of the raw frequencies.

Cohen's *d*), reasonable conformity is not perfect conformity, and it would not take much divergence to shift a result .02 standard deviations. Therefore, concluding that the brackets for Cramer's *V* more accurately represent the true difference, I would characterize the difference between Mark and referential narrative as being in the negligible to small range. I now turn to discussing the effect size results for non-referential narratives.

Table 7.3: Effect Size Measurements for Non-Referential Narrative

Subsystem	Cramer's *V* (Framework)	Cramer's *V* (Discourse)	Cohen's *d* (Framework)	Cohen's *d* (Discourse)
1	0.24 (S)	0.13 (S)	1.71 (L)	0.40 (S)
2	0.12 (S)	0.02 (N)	N/A	0.67 (M)
3	0.18 (S)	0.05 (N)	N/A	0.2 (S)
4	N/A	0.12 (S)	N/A	N/A

How Close Are the Non-Referential Narratives to Mark?

The eleven performable tests clearly show that the data from non-referential narratives match the data from Mark well. Two tests (or, potentially, three, depending on how one classifies the value of Cohen's *d* for the use of thematic subjects in projected discourse, which fell precisely on the border between negligible and small effects) indicated that the difference between them was negligible. Six (or seven) others pointed to a small effect. One each of medium and large effects round out the total. Bearing in mind the considerations mentioned in the preceding treatment of the relationship between Mark and referential narrative, the overall difference between Mark and non-referential narrative would seem to be somewhere in the range of a small effect.

Which Candidate Analogy is Closer?

The preceding subsections established that Mark's patterns of clause thematization are reasonably similar to both referential and non-referential narratives. The above discussion using the traditional brackets for interpreting Cramer's *V* and Cohen's *d* indicated that the majority of the groups of data for both referential and non-referential narratives fell into either the negligible or small category. As for which analogy is better, on the one hand, the referential comparisons had more results in the negligible category; on the other, the non-referential comparisons had one fewer medium result and an equal number of large results. The referential comparisons falling into the larger brackets tended to be closer to the lower boundary of the bracket than the non-referential comparisons in those brackets. These mixed results suggest the advisability of a more fine-grained analysis using the precise numerical results instead of the traditional brackets. The remainder of this subsection consists of this fine-grained analysis.

In the case of the frameworks' use of subsystem 1 (+/- Process), both measures of effect size agree that referential narratives are a better analogy for Mark than non-referential narratives. The brackets for Cramer's *V* showed a small difference between Mark and both referential and non-referential narratives, but comparing the actual numbers shows that the difference for non-referential narratives is almost twice as big (0.24 vs. 0.13). The corresponding results for Cohen's *d* showed a medium (0.52) and a large (1.71) divergence from Mark for referential and non-referential narratives, respectively. Thus, both measures of effect size revealed that the frameworks of referential narratives are more like Mark than the frameworks of non-referential narratives are with regard to the use of subsystem 1 (+/- Process).

The analogous comparisons for projected discourse are not as straightforward: the Cramer's *V* scores support referential narratives as the better analogy, but the Cohen's *d* scores do not. The brackets for Cramer's *V* indicated a negligible difference between Mark and referential narrative and a small difference between Mark and non-referential narrative with regard to projected discourse; comparing the numbers themselves reinforces that referential narratives (0.02) diverge from Mark far less than non-referential narratives (0.13). On the other hand, the brackets for Cohen's *d* indicated a medium divergence between Mark and referential narratives (0.58), but only a small difference between Mark and non-referential narratives (0.40). Thus, the two measures of effect size point in different directions, even allowing for how close the medium result is to the cut-off between small and medium effects, which raises the question of which measure is more reliable. In this case, Cramer's *V* seems to be the more reliable, for Cohen's *d* relies on standard deviations, and the standard deviation for this group of data from referential narrative is just over three times that of the corresponding data from Mark (28.66 selections/100 instances of the system vs. 9.44 selections/100 instances of the system for Mark), indicating problems with the reliability of the estimate of the population mean for referential narrative. Simply speaking, Cohen's d measures how many (pooled) standard deviations lie between the population means, so an unreliable estimate for one of those population means calls the reliability of Cohen's *d* into question. As a result, I conclude that Cramer's *V*, which supported referential narratives as the better analogy, is likely to be the more trustworthy measure in this case.

Turning to subsystem 2 (+/- Circumstance), the data from the frameworks of non-referential narrative did not pass the normality tests, so Cohen's *d* comparisons for the framework would be meaningless. The brackets for Cramer's *V* indicated a small divergence between Mark and both referential and non-referential narratives, though the divergence was slightly bigger in the case of referential narratives (0.16 vs. 0.12).

I was able to perform all four of the tests for the projected discourse data on the use of this subsystem, and both comparisons supported referential narrative as the better analogy, although the margin between Cramer's *V* scores was razor-thin (in both cases, 0.02 to the two decimal places I have been using).[8] The Cohen's *d* scores, however, were far enough to fall into non-neighboring brackets: the divergence between Mark and referential narratives was a negligible effect (0.07), while the divergence between Mark and non-referential narratives was a medium effect (0.67).

The samples regarding the use of subsystem 3 (+/- Subject) taken from the frameworks of both referential and non-referential narrative did not pass the normality tests, so a comparison of Cohen's *d* scores would be meaningless. Comparing the Cramer's *V* scores, however, shows a nearly identical small divergence between Mark and both referential and non-referential narratives (0.17 and 0.18, respectively).

The sample regarding the use of subsystem 3 taken from the projected discourse of non-referential narrative has a Cohen's *d* score on the cusp between a negligible and small effect (0.2). The corresponding sample from referential narrative did not pass the normality tests, however, so only Cramer's *V* is a valid measure of their relative divergence from Mark. In both cases, Cramer's *V* indicates a negligible difference between Mark and both referential and non-referential narratives, but the deviation for

[8] Without rounding, the value for referential narrative is 0.01979078, and the value for non-referential narratives is 0.0233188.

referential narratives (0.01) is smaller than the divergence for non-referential narratives (0.05).

Of the potential tests for subsystem 4 (+/- Affected), only a comparison of Cramer's *V* scores for the samples from projected discourse is appropriate. None of the samples regarding subsystem 4 passed the tests for normality, so calculating and comparing Cohen's *d* scores would be meaningless. Likewise, the samples from the framework contained too few instances for reliable χ^2 analysis, without which there is no means of calculating Cramer's *V*. The one remaining viable effect size measure indicates a negligible difference between Mark and referential narratives (0.01) and a small difference between Mark and non-referential narratives (0.12).

The measures of effect size clearly show that referential narratives are the better analogy for Mark. Eight of the ten valid comparisons support referential narrative over non-referential narrative, and one of the exceptions is a clear case of sample inadequacy (referential narrative's use of subsystem 1 in projected discourse). Although several of the differences were fairly marginal and, thus, vagaries of sampling could easily have produced a different result, the magnitude of several others is sufficient to put such concerns to rest. For instance, the Cohen's *d* score for the use of subsystem 1 (+/- Process) in the frameworks of non-referential narrative was over triple the corresponding score for referential narrative. Admittedly, that is an extreme case, but the measures of effect size indicate that non-referential narratives are more than twice as far from Mark as referential narratives on several other occasions. By contrast, in both cases where the measures of effect size indicate that referential narratives are further from Mark than non-referential narratives, the margin by which this is true is less than half, rather than double. Thus, the vast majority of the effect size comparisons place Mark closer to referential narratives than to non-referential narratives, and they do so by a greater margin than the handful of comparisons that indicated non-referential narratives were closer to Mark. This indicates that, in terms of genre, Mark belongs with referential narratives.

CHAPTER 8: SUMMARY AND SUGGESTIONS FOR ADDITIONAL RESEARCH

This project has marshalled clause thematization as a resource for providing quantitative data suggesting that histories and biographies, i.e. "referential narratives," are the proper genre analogy for the Gospel of Mark, rather than novelistic works, i.e. "non-referential narratives." The statistical tests reported in Chapter 7, particularly the effect size tests, supported this hypothesis. This final chapter looks back on the process by which I came to that conclusion, the lessons I learned along the way, and suggestions for further research I think should be done along this line. The three sections below cover these topics in the stated order.

Summary of the Project

Chapter 1 briefly surveyed literature on genre—both in general terms and with regard to Hellenistic Greek specifically—before focusing in on the history of investigation into the nature of the gospel genre in particular. The survey of this history noted that, although some have abandoned the quest for a single genre for all four Gospels, the majority position has been that they share a single genre, and the search for that genre has passed through three stages. The first stage, going at least as far back as Justin Martyr in the middle of the second century, treated the Gospels as biographies, a position virtually unchallenged until the rise of form criticism. The advent of form criticism inaugurated the second stage of discussion, during which the form-critical presupposition that the Gospels were not literature disallowed the search for parallels with the genres of contemporary literature, and the Gospels were seen as *sui generis* products of early Christian communities. Roughly speaking, this consensus held sway through the first two-thirds of the twentieth century. The crumbling of this consensus around 1970 led to the third stage of discussion in which the relationship of the Gospels to other literary genres of the ancient world was on the table once more—with an even wider variety of genres being suggested as potential analogies. The analogy taken for granted during the first stage, biography, continued to attract a great deal of attention along with its close cousin, history; perhaps most promising among the fresh analogies considered in this period was the Greco-Roman novel. Most of these investigations were qualitative and impressionistic, lacking testable criteria that would allow one to determine the relative applicability of potential analogies in an intersubjectively valid fashion. Preliminary attempts at such categories by Richard Burridge and David Mealand to buttress biographical (Burridge) and historical (Mealand) analogies fell short because neither of them convincingly answers how their chosen criteria relate to genre, stimulating me to try developing criteria that had such a link in place.

Before proceeding to develop the linguistic underpinning for my project that I hoped would set it apart from the work of Burridge and Mealand, I surveyed the literature on two related topics that arose in the course of discussing the history of the Gospel genre. The first of these is the literary level of the Greek of the New Testament as a whole and of Mark in particular. The form critics grounded their rejection of analogies on their contention that the New Testament was not literature, using its divergence from Attic as justification, and some within the third stage have grounded their particular choice of analogy (usually novels) on similar contentions. As a result, I thought it appropriate to cover the history of investigations into the relationship between the New Testament and Hellenistic Greek as a whole. After considering the possibility of the New

Testament being heavily Semitic translation Greek or a specifically Jewish dialect of Greek, this survey concluded that the Greek of the New Testament is written in Greek that very much fits within its milieu, and as such the level of language is no bar to seeking analogies within the literature of the surrounding culture. The second related topic is word order and clause structure in New Testament Greek. To some degree, this overlaps with the previous topic, because the word order of Semitic languages is one of the areas of possible areas of Semitic influence discussed in connection with the nature of New Testament Greek. Beyond this, however, clause thematization, the particular linguistic system I used as a testable criterion, has traditionally been addressed under the heading of word order, so it behooved me to lay a foundation from the previous discussion on this topic before moving on to describe how I was using it as a testable criterion.

Chapters 2 and 3 use the metaphor of cartography to illustrate the linguistic rationale for the project and to provide a structure for discussing it and the procedure for carrying it out. Essentially, in order to be useful, a map must be consistently drawn according to a particular standard, known as a projection, that reduces a three-dimensional reality to a two-dimensional picture. Cartographers have a wide variety of these to choose from, each with its own strengths and weaknesses, allowing them to choose the one that best fits the task at hand. Having decided on a particular projection, they survey the terrain, collecting the data required by the particular projection, and then draw the map according to the projection, producing a final product that represents a certain fragment of the real world in replicable terms.

The same is true of useful quantitative portraits of language use in text. Useful data on the distribution of linguistic forms results when researchers consistently apply a well-developed linguistic model. The landscape of modern linguistics offers a variety of linguistic models, each of which has its own particular strengths and weaknesses that suit it for particular tasks. Pursuing this analogy, I concluded that the ideal model for my project would have three attributes: (1) since genre is clearly a social phenomenon, it would view language in relation to its social context, (2) it would posses clear, replicable categories to ensure reliable comparisons between descriptions produced from it, and (3) it would present those descriptions in quantitative form to facilitate the statistical analyses by which I would carry out those comparisons. After considering transformational grammar and case/construction grammar as alternatives, I opted for SFL.

On that basis, I launched into a more detailed exposition of SFL. After discussing the three major dimensions through which SFL represents language, this exposition focused on the notion of register as the intersection of two of these dimensions and the relevance of this notion for my project, concluding that register is the best linguistic approximation of genre and, thus, register variation should be the best linguistic approximation of genre variation. This raised the question of how to identify register variation.

The next section explained grammatical probability and its relevance to examining register variation. With regard to the first aspect, I noted that grammatical probability, which follows directly from the notion of grammar as a closed system, holds that the observed frequencies of forms that realize systemic choice approximate the grammatical probabilities of terms in the underlying system. With regard to the second aspect, I noted that different sorts of texts, i.e. different registers, call for particular forms more than others (e.g. the imperative will be far more frequent in recipes than in a news report), so register variation would manifest as a skewing of probabilities in one or more of the systems that make up a language. This raised the question of what system or systems skew in the particular sort of texts that are the focus of my study. In the beginning, I had planned to examine three systems—one for each component of register. The search for Hellenistic Greek systems that were amenable to probabilistic analysis within the texts in which I was interested led me to focus on one in particular, the ideational theme system. The next step was to develop a system network for this system.

I wound up dividing this system into five subsystems. The first subsystem differentiated between the choice to put a finite verb at the beginning of the clause (+Process) and the choice to put anything else there (-Process). The second subsystem, which operates only if -process is selected in subsystem 1, differentiated between the choice to put an adjunct of some kind at the beginning of the clause (+Circumstance) from all other non-verb, non-adverbial choices (-Circumstance). Subsystem 3, dependent on -Process in subsystem 1 and -Circumstance in subsystem 2, differentiated between placing the subject of the clause first (+Subject) and the choice to place another sort of clause constituent typically realized by a noun (-Subject). Subsystem 4, dependent on -Subject, differentiated between the choice to place what has traditionally been called a direct object first (+Affected) from the choice to place some other sort of complement first (-Affected). Subsystem 5, the most delicate subsystem, differentiated between predicate nominatives (+Stative) and what have traditionally been called indirect objects (-Stative).

The discussion on grammatical probability left a loose end hanging, namely, how to ensure that the observed frequencies produced as close an approximation of the underlying probabilities as possible. The next section tied up this loose end by discussing corpus linguistics, i.e. the principles for deriving linguistic data from large amounts of text. The relevance of this sub-discipline for my project lay in helping me to reflect on what texts I should use to ascertain the grammatical probabilities of the terms within the ideational theme subsystem in referential and non-referential narratives. The needed reflection fell into two categories: (1) determining how many samples it would take to generate a reliable picture of referential or non-referential narratives, which I call sample depth, and (2) how long each sample needed to be in order for me to be confident it was an accurate representation of the text from which it came. The second of these was simple enough, for the literature provided a formula by which I was able to determine that—regardless of the variability of the population—a sample of 385 instances would produce an estimate of the grammatical probability that differed from the true value by no more than 5 percent. Assuming an average clause length of about twenty-five words, this led to me selecting 10,000 words as an initial goal for sample length. On the other hand, the literature did not offer a similarly clear indication relative to sample depth, so I simply decided to accumulate data on as many texts in each category as I could in the allotted time.

Chapter 3 covered the practical aspects of putting the theory contained in Chapter 2 in practice. Returning to the mapping metaphor, this would be equivalent to a cartographer telling a surveyor what places need to be surveyed and what sort of data needs to be recorded. Section 3.1 described what texts are included in each section of my comparative corpus and why some others were left out. Additionally, for each text, I include details on what edition of that text I used, whether I used a full enumeration technique or a sampling technique, and—if the latter—the rationale behind that particular sample. The next section covered two major steps in the process of generating descriptions of the texts involved in my studies. The first subsection described the process of generating box diagrams, a process that was needed for all the texts in my corpus except for Mark, which the OpenText.org project had already annotated. The second subsection described several steps that form the meat of my procedure: (1) generating a count of the absolute frequencies with which the author of a text selected each of the sorts of clause constituents that can serve as ideational theme by proceeding through the box diagrams and recording the thematization choice of each primary clause, (2) using the realization statements associated with the system network generated in Chapter 2 to transform the absolute frequencies of these realizations in text into absolute frequencies of the underlying paradigmatic choices in the various subsystems of the theme system, (3) aggregating those absolute frequencies to determine the total number of choices within each subsystem, (4) dividing the absolute frequencies for each term by the total to generate the observed proportion, which estimates the grammatical probability.

The second section described the statistical procedures for comparing the descriptions resulting from the procedures described in the first section. The first subsection described the procedures related to descriptive statistics, i.e. statistics that simply summarize a set of data. This project uses this sort of statistics in two ways: (1) the procedure for evaluating whether a sample smaller than the theoretically required size is actually reliable involves a descriptive measure of the variability of the data and (2) part of the procedure for evaluating how closely a set of data conform to the normal distribution, which is crucial for determining whether or not the t-test described below can be used on the set of data in question, involves comparing several descriptive measures. The second subsection described the procedures I used related to inferential statistics, i.e. statistical tools that allow one to evaluate how likely hypotheses are. This project used two sorts of inferential tools—χ^2 analysis and the t-test. I concluded that the former of these was better suited to my project, but I used both where the data justified use of the latter. The most important factor in determining whether I could use both tests is whether the data conformed to the normal distribution.

Chapter 4 began the work of the project in earnest by treating the thematic structure of Mark. The chapter began by explaining judgment calls I made in cases that the procedure laid out in Chapter 3 did not cover. Given the use of t-tests as an analytical tool, one important part of this chapter was to assess how closely the data collected from Mark conformed to the normal distribution. These assessments supported the use of t-tests for five subsystems: subsystems 1 and 2 in the narrative framework and subsystems 1, 2, and 3 in projected discourse.

The next section placed the quantitative data on Mark's choice of ideational theme into the qualitative SFL framework I developed in chapters 2 and 3. This process began with the observed counts of the frequencies with which Mark chose the various sorts of clause constituent as ideational theme. With the exception of Complements, which I wound up dividing into two subsystems, each of these clause constituents is the realization of one of the choices in the network of the theme system. As a result, the

"Plus Feature" column in the tables of thematization choices correlates with the appropriate frequency count in the table of realizations. The value in the "Minus Feature" column is simply the difference between the "Plus Feature" and the total for the subsystem. Another section discussed trends I observed in Mark's thematization choices. The final section of this chapter showed a distinct difference in the thematization choices of Mark's narrative framework and projected discourse.

Chapter 5 covered referential narrative. It began with a discussion of troublesome categorizations. The second section reported the descriptive statistics associated with the various groups of data associated with referential narrative. The third section assessed whether those descriptive statistics, along with other indicators like confidence intervals and boxplots, justified grouping 1 Esdras and 1 Maccabees together, as I had done on qualitative grounds, concluding that the quantitative data were consistent with the prior qualitative intuition. The next three sections consisted of tables analogous to the ones already described for Chapter 4. The penultimate section discussed several cases where the data from a unit was out of step with the trend seen in the whole document of which that unit was a part. The final section assessed the degree to which the data from referential narrative conformed to the normal distribution.

Chapter 6 continued the same tack as Chapter 5. Again, the first section covered troublesome categorization decisions. The second section reported the descriptive statistics from the three non-referential narratives. The third section assessed the validity of combining these three texts into a single category, concluding that the quantitative indicators were consistent with this qualitative judgment, although collecting additional data in the hopes of more precise estimates would be warranted. The next sections consisted of tables corresponding to the ones from Chapters 4 and 5. The penultimate section again discussed some individual units that stood out against the overall trend of their parent text. The final section compared the data from non-referential narrative to the normal distribution.

The project came to a head with Chapter 7. It compared the data on Mark (from Chapter 4) to the data on referential narrative (Chapter 5) and non-referential narrative (Chapter 6) in order to determine whether referential narrative or non-referential narrative is a closer match for Mark. The first section covered the comparisons between Mark and referential narrative; the second section covered the comparisons between Mark and non-referential narrative. Each of these sections consisted of four subsections—one each for subsystems 1 through 4. Each subsection reported the results of up to four tests: a χ^2 analysis for the narrative framework, a χ^2 analysis for projected discourse, a t-test for the narrative framework and a t-test for projected discourse. Unfortunately, a number of these tests proved to be unperformable because my data did not meet the assumptions for one or more of the tests. I found four significant deviations between Mark and referential narrative and five between Mark and non-referential narrative.

While the picture emerging from the initial hypothesis tests was fuzzier than I would have preferred, the effect size tests with which I followed up clearly supported referential narratives as the superior analogical genre for the Gospel of Mark. Effect size tests measure the divergence between two groups. For each subsystem, treating the framework portions and projected discourse sections separately, I calculated the amount by which the data accumulated for each candidate analogy diverged from the corresponding data from Mark. With those numbers in place, I then compared each effect size for referential narratives to the corresponding effect size from non-referential narrative. A smaller effect size indicated that particular group of data matched the data from Mark better than the corresponding group and, thus, the data in question pointed to its associated analogy as being a better fit for Mark. Eight of the ten performable comparisons supported referential narrative, and one of the divergent cases clearly results from problems in data collection (see the next section). Additionally, the tests supporting referential narrative often do so by a landslide, whereas the two tests supporting non-referential narrative do so by comparatively marginal amounts. Therefore, I conclude that the referential narratives are a far better fit for Mark in terms of genre than are the non-referential narratives.

The differing pictures that the hypothesis tests and measures of effect size present are at the heart of the lessons I learned in the process of conducting this project. It is to those lessons I now turn.

Lessons Learned

A project of this magnitude offers plenty of opportunities for learning things over the course of the project that would have been helpful to know from the start. This is particularly true for a project, like this one, where only limited guidance is available in the published literature and, consequently, one is to some degree forced to improvise. In the interest of furthering quantitative research into the Greek of the New Testament, sharing these lessons seems appropriate. The first group of lessons is procedural; the second group of lessons involves new vistas opened up through the use of statistical software.

Procedural Improvements

The first procedural improvement is the manner in which the data were recorded. In some ways, this is the most important one, for it is the reason I could not immediately implements some of the others when I realized they were needed. The data collection procedure I used, namely keeping a running count of the frequencies of the various options within the system I was quantifying as I moved through each sample, was adapted from the one I had used for several previous projects. There was nothing wrong with it in theory, and it had seemed practical in previous, smaller-scale projects. Lying in the woodwork, however, was a problem that became obvious once I tried to keep running frequency counts for samples as large as the texts being used in this project: each individual data point loses its identity in a running count.[1] Some of the significant effects of this became immediately obvious; others only manifested later. I took the former into account and took pains to minimize them; some of the latter surfaced late enough in the process that combating them was impractical.

As an example of a problem I could combat, I realized early on that merging the data together meant that ensuring accurate entry of the data after the fact would be difficult, if not impossible, so I took the following steps to minimize the risk: (1) I kept separate frequency counts for each chapter to make recounting easier, (2) I only started the count for a chapter when I knew I could set aside enough time to finish that chapter (or at least reach some obvious point so I could ensure that I neither skipped a primary clause nor counted one twice), (3) once I had completed the count for a chapter, I cross-checked the totals for each clause constituent type against the number of primary clauses

[1] Cf. Gries, *Statistics*, 26.

in the chapter and made sure that they were properly distributed between the framework and projected discourse on the basis of my box diagrams, and (4) I repeated the cross-checks to make sure I had done them correctly the first time. Only when I was satisfied that the data were in order did I move on. If this sounds like a time-consuming process, that is because it was, but the alternative (as far as I knew at the time) was compromising the integrity of my results. I estimate this process took at least three times as long as the improvement suggested below, which means the improvement would have allowed me to collect three times as much data in the same amount of time.

In terms of a problem I could not really combat, merging the data together in a running frequency count limited the ways I could examine the data: the format of the spreadsheet for the running count was fixed at the beginning of the project, and every clause went into one and only one cell, which meant that, for the most part, changing the data being collected would mean that I had to start over, in order to make sure the work previously done was properly integrated into the new framework. In one case, namely adding subsystem 5 (+/- Stative), doing this was practical because all the clauses that needed to be recategorized were part of the same category in the spreadsheet as it stood and the category in question was a relatively infrequent one. As it happened, the data I collected gave me no reason to reconsider the initial decisions I made in putting together the system network, other than appending subsystem 5 (+/- Stative), but had they done so, particularly at a later stage, I would have been left in an untenable position—a poor qualitative framework for my quantitative data and insufficient time to fix it. The same consideration goes double for ruling out potential nuisance variables: with the data muddled together, separating it back out to examine how a new variable affects the patterns observed is next to impossible. Although I considered this with regard to keeping separate counts for the narrative framework and projected discourse, I failed to take it to its logical extreme, i.e. every clause should have its own identity in the spreadsheet.

After adopting R, I ran across the data storage format that Stefan Gries, among many others, recommends for easily working with your data in R. As with the procedure I used, one starts with a spreadsheet file (e.g. a Microsoft Excel worksheet); the similarities end there, however. Rather than the rows representing the status of a particular clause (either narrative framework or projected discourse) and the columns designating the various realizations of choices in the ideational theme system, each row represents a clause, and each piece of information one wishes to store about that clause receives its own column.[2]

Even leaving R aside, this data format offers several advantages. First, it fixes both the problems observed above with my data entry procedure: having each observation separate means you can verify the accuracy of each row individually and rely on the computer to count them accurately, and adding extra pieces of information as a means of ruling out the influence of potential nuisance variables is as simple as adding an extra column. Second, even within the spreadsheet program itself, one can take interactions between multiple variables into account in a way that would be difficult to pull off with any conceivable table. The procedure I used was capable of handling two variables (clause status, i.e. framework/projected discourse, and type of thematic constituent). Including another variable (e.g. aspect of the clause’s main verb) would have required a

---

[2] Gries, *Statistics*, 126.

three-dimensional table for each interaction to have a cell in which I could record a frequency count, and the complexity skyrockets from there.

The ease with which this type of data can be integrated into R is simply icing on the cake. The built-in R functions to access spreadsheet data assume they will be in this format. They take the data and place it within a data structure known as a data frame. Data frames are particularly useful in R because the columns are treated as vectors, which means that any R code that takes its input as a vector can use part of a data frame (with the $ operator described in Chapter 3), but the reverse is not true: if a function requires a data frame, code that tries to give it a vector will not run. The basic statistical functions I used in Chapters 4 through 7 either require a vector (e.g. mean, sd, t.test) or at least allow vector input with some modification (e.g. chisq.test, which required me taking vectors and turning them into a simplified data frame called a matrix). Other, more complicated statistical techniques that R offers need a data frame to run (see the next subsection). With no way to get my data into this format without repeating the data collection process from scratch, I have had to forego those statistical techniques.

The second major procedural improvement relates to improving the precision of the estimates for the characteristics of referential and non-referential narratives. Chapters 5 and 6, particularly the portions comparing the various samples to verify that they formed the coherent categories suggested on qualitative grounds, repeatedly showed that the population estimates based on normalized frequencies were often at odds with those based on the raw frequencies. This seems to have resulted from at least two problems with the procedure for generating the normalized frequencies: (1) the base to which I normed the frequencies, 100 instances of the system, was too large, particularly in cases where the sample size was extremely small, and (2) I applied the normalization procedure at the chapter level, potentially exacerbating the first problem by splitting the data into more groups and, thus, decreasing the number of clauses in each group.

Choosing 100 instances of the system as the base for normalization seemed reasonable at first because each of the samples collected consists of several hundred clauses, but this fails to consider that the effective sample size for more delicate subsystems is far smaller than the total sample size. For example, if a sample has a total length of 500 primary clauses, 250 of which select +Process in subsystem 1 and 250 of which select – Process, then the effective sample size for subsystem 2 (+/- Circumstance) is only 250, for subsystem 2 depends on -Process. By the time one reaches subsystem 4 (+/- Affected) and subsystem 5 (+/- Stative), the number of choices made is almost always below ten, and sometimes substantially smaller than that. If a particular grouping of data only makes five selections within a subsystem, there are only five possible values for the normalized frequency with a base of 100 instances of the system: 0 selections/100 instances of the system (corresponding to 0 selections), 20 selections/100 instances of the system (corresponding to one selection), 40 selections/100 instances of the system (corresponding to two selections), 60 selections/100 instances of the system (corresponding to three selections), 80 selections/100 instances of the system (corresponding to four selections), and 100 selections/100 instances of the system (corresponding to five selections). When a difference of one in terms of absolute frequency equals a difference of twenty in normalized frequency, clearly the normalized frequencies are going to be unreliable; this goes double when most of the tools for which I use normalized frequencies assume that each value between the minimum and maximum are equally likely, instead of having four regions, each measuring 19 selections/100 instances of the system, that cannot occur. A normalization base of 10 instances of the system would help considerably with this problem.

Regarding the other problem, I decided to normalize the frequencies at the chapter level for practical reasons. Upon further reflection, however, this decision seems methodologically suspect. Essentially, there is no *a priori* reason to suspect that all the material in a given chapter forms a linguistically coherent unit. A related issue is that chapters vary in length, defeating the purpose of normalization, namely standardizing the units so that each data point is weighted equally. Thus, even if I found a theoretically defensible rationale for making the chapter a basic unit for analysis, I would need to apply the normalization procedure twice: the first pass would standardize the chapter-by-chapter frequencies to one another to produce an estimate for the grammatical probability of each choice in a given sample, and the second pass would standardize the estimates for each sample in one of the comparative categories (i.e. either referential or non-referential narrative) to produce an overall estimate for the grammatical probability of each choice within the comparative category as a whole. As it happens, however, R offers a better option, which I will describe in the next subsection.

Taking Better Advantage of R

The use of statistical software offers new opportunities for pushing research like this forward. The statistical procedures I followed for this project were developed at the proposal stage under the presumed constraint that the calculations needed to be ones that I felt competent to carry out by hand because the statistical software packages of which I was aware were quite expensive. R allows for more complex multivariate statistics, akin to those used in Douglas Biber's work on register variation.[3]

Besides offering other analytical tools, R also offers a better option for normalization, which is too computationally intensive to use without computer assistance.[4] This method, known as bootstrapping, essentially consists of taking a series of random samples from the data collected in order "to allow for inferences to be made about [an] unknown population from the available sample."[5] This method is gaining traction in the area of applied linguistics because it makes fewer assumptions than traditional parametric tests and allows for reliable inferences from skewed distributions.[6] One drawback is that outliers present in the original data set may happen to be overrepresented in the random sampling, but robust statistical methods that curb such a possibility exist and are recommended for use in conjunction with bootstrapping.[7] The hallmark of these robust methods is that they are based on medians, rather than means.[8]

Suggestions for Additional Research

The purpose of this section is to suggest some tasks for further research. These fall into two broad categories. The first group deals directly with the question I pursued here. The second group, on the other hand, consists of related questions that sprang up during the course of my project.

The issue of corpus size raises its head in two ways: (1) each sample needs to be sufficiently large to be a reliable picture of the document from which it is taken and 2)

[3] Cf. Arnold et al., "Beyond Lexical Frequencies," 718–19.
[4] Nikitina and Furuoka, "Expanding the Methodological Arsenal," 422.
[5] Nikitina and Furuoka, "Expanding the Methodological Arsenal," 422.
[6] Nikitina and Furuoka, "Expanding the Methodological Arsenal," 422.
[7] Nikitina and Furuoka, "Expanding the Methodological Arsenal," 423.
[8] Nikitina and Furuoka, "Expanding the Methodological Arsenal," 423.

the various components of the corpus—referential and non-referential narrative—need to consist of a sufficient number of samples to be reliable pictures of the candidate analogies they purport to represent. With regard to the first issue, as discussed in Chapter 3 above, statistical theory indicated that 385 units—primary clauses in this case—would be a sufficient sample size to generate a reliable picture of the document as a whole. With regard to the second, I found no such external guidance, but I would suggest that the proof is ultimately in the pudding: when the data begin clustering, and thus a clear picture of the underlying system emerges, then one has accumulated enough instances to have a reliable picture of the system. With these twin issues of sample length and sample depth in mind, one can clearly see that further quantitative research into Hellenistic Greek grammar, and the issue of the gospel genre in particular, would benefit from greatly enlarging the comparative corpus.

Besides needing to enlarge the corpus, another way of perhaps turning up data more suited to answering my primary research question is improving the system network I used as the basis for grammatical probabilities. In particular, I am aware of the lack of straightforward realization statements for subsystem 2. Whereas one choice within other subsystems factors out one identifiable realization and the other choice simply opens up the next subsystem, the choice +Circumstance has several different realizations.

In terms of additional questions raised during the course of research, perhaps the most significant is that the data I have collected, such as it is, seems to indicate that clause thematization more closely relates to the type of discourse (narration vs. reported speeches) within narrative texts than to what kind of narrative a particular text is (i.e.

referential narrative vs. non-referential narratives).[9] This raises interesting questions regarding the narrative techniques that authors deploy in Hellenistic Greek narratives. Some of these are difficult to answer without access to living native speakers, such as adjudicating whether the patterns of referential or non-referential narratives more accurately represent the patterns of spoken language. Reformatting the data into the long form described in the preceding section would allow the use of an ANOVA test to quantify how much of the variation the grouping of texts explains versus the amount of variation explained by a particular clause's status as part of either the framework or projected discourse.[10]

Lastly, bearing in mind the concerns relative to sample size above, my findings do not all support Halliday's notion that systems will be either roughly equiprobable (0.5/0.5) or maximally skewed (0.9/0.1).[11] The framework of Mark shows quite a bit of correspondence with Halliday's prediction: in this group the probabilities of all five subsystems differ from one of Halliday's predictions by no more than 5 percent (the maximum deviations are 0.55/0.45 for subsystems 1 and 2 and 0.85/0.15 for subsystem 4). Mark's projected discourse tells a different story, with only the profile of subsystem 5 falling within these bounds (0.86/0.14). In 1 Maccabees, only subsystem 5 in the framework (0.5/0.5, but, importantly, only two instances) and subsystem 1 in discourse (0.51/0.49) deviate from Halliday's prediction by no more than 5 percent. In 1 Esdras, subsystem 1 in the framework (0.49/0.51) and subsystem 3 in discourse (0.54/0.46) fit the

---

[9] The role that a particular clause's status as part of either the framework or discourse played in determining the sort of clause constituent that would appear as ideational theme thus proved to be a "confounding variable," which "are the most pesky because we often do not know they exist, or we find out about their existence after the data have been gathered and analyzed" (Eddington, *Statistics*, 9).

[10] Cf. Manilla et al., "Quantifying Variation," 342.

[11] E.g. Halliday, "System and Instance," 81.

bill, though a few others were close. In Judith, only subsystem 5 in discourse (0.5/0.5, but once again only two instances) closely matches one of Halliday's predicted profiles for systems. Both subsystems 2 (0.45/0.55) and 3 (0.85/0.15) of Tobit's framework are close to one of Halliday's predictions, as are subsystems 1 (0.52/0.48) and 3 (0.5/0.5) in discourse. Lastly, the framework of Joseph and Aseneth has two instances of correspondence—subsystem 1 (0.89/0.11) and subsystem 3 (0.85/0.15)—and subsystem 3 is the only case in discourse (0.87/0.13). Ten sets of data per text and 6 texts equals 60 chances for correspondence, yet I observed only 18 instances of correspondence—a ratio of 30 percent, with many being either borderline or based on miniscule sample sizes.

The last-minute addition of skewness and kurtosis as descriptive statistics uncovered an interesting pattern. The framework data and projected discourse data from Mark skew in the same direction only once, namely both groups exhibit negative skew in subsystem 3. For subsystems 1 and 2, the framework data have negative skew and the discourse data have positive skew; for subsystems 4 and 5, the framework data have positive skew and the discourse data have negative skew. To borrow language that will be familiar, the pattern here is chiastic, with subsystem 3 in the middle.

In summary, further studies need to consider what sources of variation they choose to investigate and the systems in which they investigate that variation. Such studies would do well not to assume that systems in Hellenistic Greek match the dual profile Halliday posits, where grammatical probabilities have only two profiles—roughly equiprobable (0.5/0.5) or radically skew (0.9/0.1). I hope these findings have shown the potential of quantitative grammatical investigation in the study of the Greek of the New Testament and Hellenistic Greek as a whole.

BIBLIOGRAPHY

Adams, Sean A. "Atticism, Classicism, and Luke–Acts: Discussions with Albert Wifstrand and Loveday Alexander." In *The Language of the New Testament: Context, History, and Development*, edited by Stanley E. Porter and Andrew W. Pitts, 91–111. LBS 6. Leiden: Brill, 2013.

———. *The Genre of Acts and Collected Biography*. SNTSMS 156. Cambridge: Cambridge University Press, 2013.

Alexander, C. Loveday. "Luke's Preface in the Pattern of Greek Preface-Writing." *NovT* 28 (1986) 48–74.

Arnold, Taylor, and Lauren Tilton. *Humanities Data in R: Exploring Networks, Geospatial Data, Images, and Text*. QMHSS. New York: Springer International, 2015.

Arnold, Taylor, et al. "Beyond Lexical Frequencies: Using R for Text Analysis in the Digital Humanities." *Language Resources & Evaluation* 53 (2019) 707–33.

Aune, David E. "Genre Theory and the Genre Function of Mark and Matthew." In *Mark and Matthew I, Comparative Readings: Understanding the Earliest Gospels in Their First-Century Settings*, edited Eve-Marie Becker and Anders Runesson, 145–75. WUNT 271. Tübingen: Mohr Siebeck, 2011.

———. "Greco-Roman Biography." In *Greco-Roman Literature and the New Testament*, edited by David E. Aune, 107–26. SBLSBS 21. Atlanta: Scholars, 1988.

Bartlett, John R. "1 Maccabees." In *Eerdmans Commentary on the Bible*, edited by James D. G. Dunn and John W. Rogerson, 807–30. Grand Rapids: Eerdmans, 2003.

Becker, Eve-Marie. "The Gospel of Mark in the Context of Ancient Historiography." In *The Function of Ancient Historiography in Biblical and Cognate Studies*, edited by Patricia G. Kilpatrick and Timothy Goltz, 124–35. LHBOTS 489. New York: T. & T. Clark, 2008.

Biber, Douglas A., and Jessie Egbert. *Register Variation Online*. Cambridge: Cambridge University Press, 2018.

Bird, Michael F. *1 Esdras: Introduction and Commentary on the Greek Text in Codex Vaticanus*. Septuagint Commentary Series. Leiden: Brill, 2012.

Black, Stephanie L. *Sentence Conjunctions in the Gospel of Matthew: καί, δέ, γάρ, οὖν, and Asyndeton in Narrative Discourse*. JSNTSup 216. SNTG 9. Sheffield: Sheffield Academic, 2002.

Bloomfield, Leonard. *Language*. 1933. Reprint, Chicago: University of Chicago Press, 1984.

Breytenbach, Cilliers. “Current Research on the Gospel of Mark: A Report on Monographs Published from 2000–2009.” In *Mark and Matthew I, Comparative Readings: Understanding the Earliest Gospels in Their First-Century Settings*, edited Eve-Marie Becker and Anders Runesson, 13–32. WUNT 271. Tübingen: Mohr Siebeck, 2011.

———. *Nachfolge und Zukunftserwartung nach Markus: Eine methodenkritische Studie*. ATANT 71. Zürich: Theologischer Verlag, 1984.

Buchstaller, Isabelle, and Ghada Khattab. “Population Samples.” In *Research Methods in Linguistics*, edited by Robert J. Podesva and Devyani Sharma, 74–95. Cambridge: Cambridge Univeristy Press, 2013.

Bultmann, Rudolf. *The History of the Synoptic Tradition*. Translated by John Marsh. Second edition. Oxford: Basil Blackwell, 1968.

Burchard, C. “Joseph and Aseneth.” In *OTP* 2:177–201.

Burney, C. F. *Aramaic Origin of the Fourth Gospel*. Oxford: Clarendon, 1922.

Burridge, Richard A. “About People, by People, for People: Gospel Genre and Audiences.” In *The Gospels for All Christians*, edited by Richard Bauckham, 113–41. Grand Rapids: Eerdmans, 1998.

———. *What Are the Gospels? A Comparison with Greco-Roman Biography*. Second edition. The Biblical Resource Series. Grand Rapids: Eerdmans, 2004.

Burrows, Millar. “Mark’s Transitions and the Translation Hypothesis.” *JBL* 48 (1929) 117–23.

Butler, Christopher. *Statistics in Linguistics*. Oxford: Basil Blackwell, 1985.

Cancik, Hubert. “*Bios und Logos*: Formengeschichtliche Untersuchungen zu Lukians >Demonax<.” In *Markus-Philologie: historische, literaturgeschichctliche und stylistische Untersuchungen zum zweiten Evangelium*, edited by Hubert Cancik, 114–29. WUNT 33. Tübingen: Mohr-Siebeck, 1984.

———. “Die Gattung Evangelium: Das Markus-Evangelium im Rahmen der antiken Historiographie.” In *Markus-Philologie: historische, literaturgeschichtliche, und stylistische Untersuchungen zum zweiten Evangelium*, edited by Hubert Cancik, 85–112. WUNT 33. Tübingen: Mohr–Siebeck, 1984.

Cantos, Pascual. “The Use of Linguistic Corpora for the Study of Linguistic Variation and Change: Types and Computational Applications.” In *The Handbook of Historical Sociolinguistics*, edited by Juan M. Hernández–Campoy and Juan C. Conde-Silvestre, 99–122. BHL. Chichester: Wiley-Blackwell, 2012.

Chambers, J. K. “Studying Language Variation: An Informal Epistemology.” In *The Handbook of Language Variation and Change*, edited by J.K. Chambers et al., 1–15. 2nd ed. BHL. Chichester: Wiley-Blackwell, 2013.

Cook, John G. *The Structure and Persuasive Power of Mark*. SemeiaSt 28. Atlanta: Scholars Press, 1995.

Crellin, Robert Samuel David. *The Syntax and Semantics of the Perfect Active in Literary Koine Greek*. Publications of the Philological Society 47. Malden, MA: Wiley-Blackwell, 2016.

Danove, Paul L. *The End of Mark’s Story: A Methodological Study*. BibInt 3. Leiden: Brill, 1993.

———. *Linguistics and Exegesis in the Gospel of Mark: Applications of a Case Frame Analysis and Lexicon*. JSNTSup 218. SNTG 10. Sheffield: Sheffield Academic, 2001.

———. *The Rhetoric of Characterization of God, Jesus, and Jesus’ Disciples in the Gospel of Mark*. JSNTSup 290. New York: T. & T. Clark, 2005.

———. “The Theory of Construction Grammar and its Application to Biblical Greek.” In *Biblical Greek Language and Linguistics: Open Questions in Current Research*, edited by Stanley E. Porter and D. A. Carson, 119–50. JSNTSup 80. Sheffield: JSOT Press, 1993.

———. “Verbs of Experience: Toward a Lexicon Detailing the Argument Structures Assigned by Verbs.” In *Linguistics and the New Testament: Critical Junctures*, edited by Stanley E. Porter and D.A. Carson, 144–205. JSNTSup 168. SNTG 5. Sheffield: Sheffield Academic, 1999.

Dawson, Zachary K. “The Problem of Gospel Genres: Unmasking a Flawed Consensus and Providing a Fresh Way Forward with Systemic Functional Linguistics Genre Theory.” *BAGL* 8 (2019) 33–77.

Decker, Rodney J. “Markan Idiolect in the Study of the Greek of the New Testament.” In *The Language of the New Testament: Context, History, and Development*, edited by Stanley E. Porter and Andrew W. Pitts, 43–66. LBS 6. Leiden: Brill, 2013.

Deissmann, Adolf. “Hellenistic Greek.” In *The Language of the New Testament: Classic Essays*, edited by Stanley E. Porter, 39–59. Translated by Marika Walter, Harold Biessmann, and Stanley E. Porter. JSNTSup 60. Sheffield: JSOT Press, 1991.

Desagulier, Guillaume. *Corpus Linguistics and Statistics with R: Introduction to Quantitative Methods in Linguistics*. QMHSS. New York: Springer International, 2017.

Dibelius, Martin. *From Tradition to Gospel*. Translated by Bertram Lee Woolf. 1935. Reprint, Cambridge: James Clarke, 1982.

Diehl, Judith. “What is a ‘Gospel’? Recent Studies in the Gospel Genre.” *CurBR* 9 (2011) 171–99.

Dik, Helma. *Word Order in Ancient Greek: A Pragmatic Account of Word Order Variation in Herodotus*. Amsterdam Studies in Classical Philology 5. Amsterdam: J.C. Gieben, 1995.

Dormeyer, Detlev. *The New Testament among the Writings of Antiquity*, edited by Stanley E. Porter. Translated by Rosemarie Kossov. Sheffield: Sheffield Academic, 1998.

———. “Plutarchs Cäsar und die Erste Evangeliumsbiographie des Markus.” In *Rom und das himmlische Jerusalem: Die frähen Christen zwischen Anpassung und Ablehung*, edited by Raban von Haehling, 29–52. Darmstadt: Wissenschaftliche Buchgesellschaft, 2000.

Duff, David. “Introduction.” In *Modern Genre Theory,* edited by David Duff, 1–24. Longman Critical Readers. London: Longman, 2000.

———. “Key Concepts.” In *Modern Genre Theory*, edited by David Duff, x–xvi. Longman Critical Readers. London: Longman, 2000.

Duling, D. C. “Testament of Solomon: A New Translation and Introduction.” In *OTP* 1:935–87.

Dunn, Graham. “Syntactic Word Order in Herodotean Greek.” *Glotta* 66 (1988) 63–79.

Dvorak, James D. “Martin Dibelius and Rudolf Bultmann.” In *Prevailing Methods before 1980,* edited by Stanley E. Porter and Sean A. Adams, 257–77. Vol. 1 of *Pillars in the History of Biblical Interpretation*. McMaster Biblical Studies Series 2. Eugene, OR: Pickwick, 2016.

Dvorak, James D., and Ryder Dale Walton. “Clause as Message: Theme, Topic, and Information Flow in Mark 2:1–12 and Jude.” *BAGL* 3 (2014) 31–85.

Eddington, David. *Statistics for Linguistics: A Step-by-Step Guide for Novices*. Newcastle upon Tyne: Cambridge Scholars, 2015.

Fillmore, Charles J. “The Case for Case.” In *Universals in Linguistic Theory,* edited by Emmon Bach and Robert T. Harms, 1–88. New York: Holt, Rinehart, and Wilson, 1968.

Fishelov, David. *Metaphors of Genre: The Role of Analogies in Genre Theory*. University Park, PA: Pennsylvania State University Press, 1993.

Goodspeed, Edgar J. *New Chapters in New Testament Study*. New York: Macmillan, 1937.

———. “The Origin of Acts.” *JBL* 39 (1920) 83–101.

Gray, Bethany, and Douglas Biber. “Corpus Approaches.” In *Continuum Companion to Discourse Analysis*, edited by Ken Hyland and Brian Paltridge, 138–52. London: Continuum, 2011.

Gries, Stefan Th. “Basic Significance Testing.” In *Research Methods in Linguistics*, edited by Robert J. Podesva and Devyani Sharma, 257–81. Cambridge: Cambridge University Press, 2014.

———. Elementary Statistical Testing with R.” in *Research Methods in Language Variation and Change*, edited by Manfred Krug and Julia Schlüter, 361–81. Cambridge: Cambridge University Press, 2013.

———. “Sources of Variability Relevant to the Cognitive Sociolinguist, and Corpus- as well as Psycholinguistic Methods and Notions to Handle Them.” *Journal of Pragmatics* 52 (2013) 5–16.

———. *Statistics for Linguistics with R: A Practical Introduction*. 2nd edition. Mouton Textbook Series. Berlin: de Gruyter, 2013.

Gries, Stefan Th., and Nick C. Ellis. “Statistical Measures for Usage-Based Linguistics.” *Language Learning* 65 (Supplement 1 2015) 228–55.

Gries, Stefan Th., and John Newman. “Creating and Using Corpora.” In *Research Methods in Linguistics*, edited by Robert J. Podesva and Devyani Sharma, 257–81. Cambridge: Cambridge University Press, 2014.

Guelich, Robert A. “The Gospel Genre.” In *The Gospel and the Gospels*, edited by Peter Stuhlmacher, 173–208. Grand Rapids: Eerdmans, 1991.

Gundry, Robert H. “The Symbiosis of Theology and Genre Criticism of the Canonical Gospels.” In *The Old is Better: New Testament Essays in Support of Traditional Interpretations*, 18–48. WUNT 178. Tübingen: Mohr Siebeck, 2011.

Güttgemanns, Erhardt. *Candid Questions Concerning Gospel Form Criticism: A Methodological Sketch of the Fundamental Problematics of Form and Redaction Criticism*. Translated by William G. Doty. PTMS 26. Pittsburgh: Pickwick, 1979.

Hägg, Tomas. *The Art of Biography in Ancient Antiquity*. Cambridge: Cambridge University Press, 2012.

Hale, Ken. “Walpiri and the Grammar of Non-Configurational Languages.” *Natural Languages and Linguistic Theory* 1 (1983) 5–47.

Halliday, M. A. K. “Appendix: Systemic Theory.” In *On Language and Linguistics*, edited by Jonathan J. Webster, 433–41. CWH 3. London: Continuum, 2003.

———. “Categories of the Theory of Grammar.” In *On Grammar*, edited by Jonathan J. Webster, 37–94. CWH 1. London: Continuum, 2002.

———. “Corpus Studies and Probabilistic Grammar.” In *Computational and Quantitative Studies*, edited by Jonathan J. Webster, 63–75. CWH 6. London: Continuum, 2005.

———. “The Gloosy Ganoderm: Systemic Functional Linguistics and Translation.” In *Halliday in the 21st Century*, edited by Jonathan J. Webster, 5–34. CWH 11. London: Bloomsbury Academic, 2013.

———. *An Introduction to Functional Grammar*. London: Charles Arnold, 1985.

———. *Language as Social Semiotic: The Social Interpretation of Language and Meaning*. London: Edward Arnold, 1978.

———. “Language as System and Language as Instance.” In *Computational and Quantitative Studies*, edited by Jonathan J. Webster, 76–92. CWH 6. London: Continuum, 2005.

———. “Linguistic Function and Literary Style: An Inquiry into the Language of William Golding’s *The Inheritors*.” In *Linguistic Studies of Text and Discourse*, edited by Jonathan J. Webster, 88–125. CWH 2. London: Continuum, 2005.

———. “Linguistics and Machine Translation.” In *Computational and Quantitative Studies*, edited by Jonathan J. Webster, 20–36. CWH 6. London: Continuum, 2005.

———. “On Grammar as the Driving Force from Primary to Higher-Order Consciousness.” In *Halliday in the 21st Century*, edited by Jonathan J. Webster, 159–89. CWH 11. London: Bloomsbury Academic, 2013.

———. “On Language in Relation to Fuzzy Logic and Intelligent Computing.” In *Computational and Quantitative Studies*, edited by Jonathan J. Webster, 196–212. CWH 6. London: Continuum, 2005.

———. “On Text, and Discourse, Information, and Meaning.” In *Halliday in the 21st Century*, edited by Jonathan J. Webster, 55–70. CWH 11. London: Bloomsbury Academic, 2013.

———. “Pinpointing the Choice: Meaning and the Search for Equivalents in a Translated Text.” In *Halliday in the 21st Century*, edited by Jonathan J. Webster, 143–51. CWH 11. London: Bloomsbury Academic, 2013.

———. “Putting Linguistic Theory to Work.” In *Halliday in the 21st Century*, edited by Jonathan J. Webster, 127–41. CWH 11. London: Bloomsbury Academic, 2013.

———. “Quantitative Studies and Probabilities in Grammar.” In *Computational and Quantitative Studies*, edited by Jonathan J. Webster, 130–56. CWH 6. London: Continuum, 2002.

———. “The Spoken Language Corpus: A Foundation for Grammatical Theory.” In *Halliday in the 21st Century*, edited by Jonathan J. Webster, 5–34. CWH 11. London: Bloomsbury Academic, 2013.

———. “Text as Semantic Choice in Social Contexts.” In *Linguistic Studies of Text and Discourse*, edited by Jonathan J. Webster, 23–81. CWH 2. London: Continuum, 2002.

———. “Towards Probabilistic Interpretations.” In *Computational and Quantitative Studies*, edited by Jonathan J. Webster, 42–62. CWH 6. London: Continuum, 2005.

———. “Why Do We Need to Understand Language?” In *Halliday in the 21st Century*, edited by Jonathan J. Webster, 71–81. CWH 11. London: Bloomsbury Academic, 2013.

———. “Working with Meaning: Towards and Appliable Linguistics.” In *Halliday in the 21st Century*, edited by Jonathan J. Webster, 35–54. CWH 11. London: Bloomsbury Academic, 2013.

Halliday, M.A.K., and Ruqaiya Hasan. *Language, Context, and Text: Aspects of Language in a Social-Semiotic Perspective.* 2nd ed. Oxford: Oxford University Press, 1989.

Halliday, M.A.K., and Z.L. James. “A Quantitative Study of Polarity and Primary Tense in the English Finite Clause.” In *Computational and Quantitative Studies*, edited by Jonathan J. Webster, 93–129. The CWH 6. London: Continuum, 2005.

Halliday, M.A.K., and Jonathan J. Webster. *Text Linguistics: The How and Why of Meaning*. Sheffield: Equinox, 2014.

Hasan, Ruqaiya. “Language and Society in a Systemic-Functional Perspective.” In *Context in the System and Process of Language*, 5–33. The Collected Works of Ruqaiya Hasan 4. Sheffield: Equinox, 2016.

———. “Meaning, Context, and Text: Fifty Years after Malinowski.” In *Context in the System and Process of Language*, 34–77. The Collected Works of Ruqaiya Hasan 4. Sheffield: Equinox, 2016.

———. “‘What’s Going On?’: A Dynamic View of Context in Language.” In *Context in the System and Process of Language*, 78–94. The Collected Works of Ruqaiya Hasan 4. Sheffield: Equinox, 2016.

He, Qingshun, and Bingjun Yang. *Absolute Clauses in English from the Systemic Functional Perspective: A Corpus-Based Study*. The M.A.K. Halliday Library Functional Linguistics Series. Berlin: Springer, 2015.

Hernández-Campoy, Juan M., and Natalie Schilling. “The Application of the Quantitative Paradigm to Historical Sociolinguistics: Problems with the Generalizability Principle.” In

The Handbook of Historical Sociolinguistics, edited by Juan M. Hernández-Campoy and Juan Camilo Conde-Silvestre, 63–75. BHL. Chichester: Wiley-Blackwell, 2012.

Hock, Ronald F. "The Greek Novel." In *Greco-Roman Literature and the New Testament*, edited by David E. Aune, 127–46. SBLSBS 21. Atlanta: Scholars, 1988.

Holzberg, Niklas. *The Ancient Novel: An Introduction*. Translated by Christine Jackson-Holzberg. London: Routledge, 1995.

Isaac, E. "1 (Ethiopic Apocalypse of) Enoch: A New Translation and Introduction." In *OTP* 1:5–89.

Jenset, Gard B., and Barbara McGillivray. *Quantitative Historical Linguistics: A Corpus Framework*. Oxford Studies in Diachronic and Historical Linguistics 26. Oxford: Oxford University Press, 2017.

Jockers, Matthew. *Text Analysis with R for Students of Literature*. QMHSS. New York: Springer International, 2014.

Johnson, Daniel Ezra. "Descriptive Statistics." In *Research Methods in Linguistics*, edited by Robert J. Podesva and Devyani Sharma, 288–315. Cambridge: Cambridge University Press, 2014.

Johnson, Keith. *Quantitative Methods in Linguistics*. Oxford: Blackwell, 2008.

Kendall, Tyler. "Data in the Study of Variation and Change." In *The Handbook of Language Variation and Change*, edited by J.K. Chambers et al., 38–56. 2nd ed. BHL. Chichester: Wiley-Blackwell, 2013.

Kilpatrick, G.D. "Some Notes on Marcan Usage." In *The Language and Style of the Gospel of Mark*, edited by J.K. Elliott, 159–74. NovTSup 71. Leiden: Brill, 1993.

Köstenberger, Andreas, et al. *Going Deeper with New Testament Greek: An Intermediate Study of the Grammar and Syntax of the New Testament*. Nashville, TN: Broadman & Holman Academic, 2016.

Kwong, Ivan Shing Chung. *The Word Order of the Gospel of Luke: Its Foregrounded Messages*. LNTS 298. Edinburgh: T. & T. Clark, 2005.

Land, Christopher D. "Varieties of the Greek Language." In *The Language of the New Testament: Context, History, and Development*, edited by Stanley E. Porter and Andrew W. Pitts, 243–60. LBS 6. Leiden: Brill, 2013.

Land, Christopher D., and Francis G. H. Pang. "The Past, Present, and Future of the OpenText.org Annotated Greek Corpus." In *The Language and Literature of the New*

Testament: Essays in Honor of Stanley E. Porter's 60th Birthday, edited by Lois K. Dow et al., 69–105. BibInt 150. Leiden: Brill, 2016.

Larson-Hall, Jenifer, and Luke Plonsky. "Reporting and Interpreting Quantitative Research Findings: What Gets Reported and Recommendations for the Field." *Language Learning* 65 (Supplement 1 2015) 127–59.

Lee, John A. L. "The Atticist Grammarians." In *The Language of the New Testament: Context, History, and Development*, edited by Stanley E. Porter and Andrew W. Pitts, 283–308. LBS 6. Leiden: Brill, 2013.

Levinsohn, Steven H. "Discourse Study of Constituent Order and the Article in Philippians." In *Discourse Analysis and Other Topics in Biblical Greek*, edited by Stanley E. Porter and D. A. Carson, 69–74. JSNTSup 113. Sheffield: Sheffield Academic, 1995.

Libby, James A. "Disentangling Authorship and Genre in the Greek New Testament: History, Method, and Praxis." PhD diss., McMaster Divinity College, 2015.

Lyons, John. *An Introduction to Theoretical Linguistics*. Cambridge: Cambridge Univeristy Press, 1968.

Maloney, Elliot C. *Semitic Interference in Markan Syntax*. SBLDS 51. Chico, CA: Scholars Press, 1981.

Mannila, Heikki, et al. "Quantifying Variation and Estimating the Effects of Sample Size on the Frequencies of Linguistic Variables." In *Research Methods in Language Variation and Change*, edited by Manfred Krug and Julia Schlüter, 337–60. Cambridge: Cambridge University Press, 2013.

Martin, James R. *English Text: Systems and Structure*. Philadelphia: John Benjamins, 1992.

Martin, James R., and Ruth Wodak. "Introduction." In *Re/reading the Past: Critical and Functional Perspectives on Time and Value*, edited by James R. Martin and Ruth Wodak, 1–16. Discourse Approaches to Politics, Society, and Culture 8. Amsterdam: Benjamins, 2003.

Martín-Asensio, Gustavo. *Transitivity-Based Foregrounding in the Acts of the Apostles: A Functional-Grammatical Approach to the Lukan Perspective*. JSNTSup 202. SNTG 8. Sheffield: Sheffield Academic, 2000.

Mathewson, David L., and Elodie Ballantine Emig. *Intermediate Greek Grammar: Syntax for Students of the New Testament*. Grand Rapids: Baker Academic, 2016. Kindle edition.

Matthiessen, Christian M. I. M. *Lexicogrammatical Cartography: English Systems*. Textbook Series in the Language Sciences. Tokyo: International Language Science Publishers, 1995.

Matthiessen, Christian M. I. M., et al., eds. *Key Terms in Systemic Functional Linguistics*. Key Terms Series. London: Continuum, 2010.

McCarthy, Michael, and Anne O'Keeffe. "Historical Perspective: What are Corpora, and How Have They Evolved?" In *The Routledge Handbook of Corpus Linguistics*, edited by Anne O'Keeffe and Michael McCarthy, 3–13. Routledge Handbooks in Applied Linguistics. London: Routledge, 2010.

Mealand, David L. "Computers in New Testament Research: An Interim Report." *JSNT* 33 (1988) 97–115.

———. "Hellenistic Greek and the New Testament: A Stylometric Perspective." *JSNT* 34 (2012) 323–45.

———. "Hellenistic Historians and the Style of Acts." *ZNW* 82 (1991) 42–66.

Momigliano, Arnaldo. *The Development of Greek Biography*. Expanded edition. Cambridge, MA: Harvard University Press, 1993.

Montgomery, James A. *The Origin of the Gospel According to St. John*. Philadelphia: John C. Winston, 1923.

Morgan, J. R. "Fiction and History: Historiography and the Novel." In *A Companion to Greco-Roman Historiography*, edited by John Marincola, 2:553–64. Oxford: Blackwell, 2007.

Moule, C. F. D. *An Idiom-Book of New Testament Greek*. 2nd edition. Cambridge: Cambridge University Press, 1959.

Moulton, James Hope. "Albert Thumb." *Classical Review* 29 (1915) 221–22.

———. "New Testament Greek in the Light of Modern Discovery." In *The Language of the New Testament: Classic Essays*, edited by Stanley E. Porter, 60–97. JSNTSup 60. Sheffield: JSOT Press, 1991.

Nevalainen, Terttu, and Helena Raumolin-Brunberg. "Historical Sociolinguistics: Origins, Motivations, and Paradigms." In *The Handbook of Historical Sociolinguistics*, edited by Juan Manuel Hernández-Campoy and Juan Camilo Conde–Silvestre, 22–40. BHL. Chichester: Wiley-Blackwell, 2012.

Nikitina, Larisa, and Fumitaka Furuoka. "Expanding the Methodological Arsenal of Applied Linguistics with a Robust Statistical Procedure." *Applied Linguistics* 39 (2018) 422–28.

Norden, Eduard. *Die Antike Kunstprosa: Vom VI. Jahrhundert v. Chr. bis in die Zeit der Renaissance*. 2 vols. [n.d]. Reprint, Stuttgart: B.G. Teubner, 1958.

Norris, John M. “Statistical Significance Testing in Second Language Research: Basic Problems and Suggestions for Reform.” *Language Learning* 65 S1 (2015) 97–126.

Norris, John M., et al. “Improving Second Language Quantitative Research.” *Language Learning* 65 S1 (2015) 1–8.

Nylund, Jan H. “The Prague School of Linguistics and Its Influence on New Testament Language Studies.” In *The Language of the New Testament: Context, History, and Development*, edited by Stanley E. Porter and Andrew W. Pitts, 155–21. LBS 6. Leiden: Brill, 2013.

O’Donnell, Matthew B. *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, 206–62. JSNTSup 168. SNTG 5. Sheffield: Sheffield Academic, 1999.

O’Donnell, Matthew Brook, et al., eds. “Clausal Level Annotation Specification Version 0.2,” July, 7, 2004, http://openttext.org/model/guidelines/clause/0-2.html.

Palmer, Michael W. “How Do We Know a Phrase is a Phrase? A Plea for Procedural Clarity in the Application of Linguistics to Biblical Greek.” In *Biblical Greek Language and Linguistics: Open Questions in Current Research*, edited by Stanley E. Porter and D.A. Carson, 152–86. JSNTSup 80. Sheffield: JSOT Press, 1993.

———. *Levels of Constituent Structure in New Testament Greek*. SBG 4. New York: Peter Lang, 1995.

Pearson, Brook W. R., and Stanley E. Porter. “The Genres of the New Testament.” In *A Handbook to the Exegesis of the New Testament*, edited by Stanley E. Porter, 131–65. NTTS 25. Leiden: Brill, 1997.

Petersen, Norman R. “Can One Speak of a Gospel Genre?” *Neot* 28 (1994) 137–158.

Pitts, Andrew W. “The Genre of the Third Gospel and Authoritative Citation.” PhD dissertation, McMaster Divinity College, 2014.

———. “Word Order and Clause Structure: A Comparative Study of Some New Testament Corpora.” In *The Language of the New Testament: Context, History, and Development*, edited by Stanley E. Porter and Andrew W. Pitts, 311–46. LBS 6. Leiden: Brill, 2013.

Pitts, Andrew W., and Joshua D. Tyra. “Exploring Linguistic Variation in an Ancient Greek Single-Author Corpus: A Register Design Analysis of Josephus and Pauline Pseudonymity.” In *The Language and Literature of the New Testament: Essays in Honor* *of Stanley E. Porter's 60th Birthday*, edited by Lois K. Dow, 257–83. BibInt 150. Leiden: Brill, 2017.

Porter, Stanley E. "Dialect and Register in the Greek of the New Testament: Theory." In *Rethinking Contexts, Rereading Texts: Contributions from the Social Sciences to Biblical Interpretation*, edited by M. Daniel Carroll R., 190–208. JSOTSup 299. Sheffield: Sheffield Academic, 2000.

———. "The Greek of the New Testament as a Disputed Area of Research." In *The Language of the New Testament: Classic Essays*, edited by Stanley E. Porter, 11–38. JSNTSup 60. Sheffield: JSOT Press, 1991.

———. "Greek Word Order: Still an Unexplored Area in New Testament Studies?" In *Linguistic Analysis of the Greek New Testament: Studies in Tools, Methods, and Practice*, 347–62. Grand Rapids: Baker Academic, 2015.

———. "The Ideational Metafunction and Register." In *Linguistic Analysis of the Greek New Testament: Studies in Tools, Methods, and Practice*, 145–58. Grand Rapids: Baker Academic, 2015.

———. *Idioms of the Greek New Testament*. 2nd edition. Biblical Languages: Greek 2. Sheffield: Sheffield Academic, 1994.

———. "An Introduction to Other Topics in Biblical Greek Language and Linguistics." In *Biblical Greek Language and Linguistics: Open Questions in Current Research*, edited by Stanley E. Porter and D. A. Carson, 84–89. JSNTSup 80. Sheffield: JSOT Press, 1993.

———. "Matthew and Mark: The Contribution of Recent Linguistic Thought." In *Mark and Matthew I, Comparative Readings: Understanding the Earliest Gospels in their First-Century Settings*, edited by Eve-Marie Becker and Anders Runesson, 97–119. WUNT 271. Tübingen: Mohr Siebeck, 2011.

———. "A Multidisciplinary Approach to Exegesis." In *Linguistic Analysis of the Greek New Testament: Studies in Tools, Methods, and Practice*, 93–112. Grand Rapids: Baker Academic, 2015.

———. "A Register Analysis of Mark 13: Towards a Context of Situation." In *Linguistic Analysis of the Greek New Testament: Studies in Tools, Methods, and Practice*, 219–36. Grand Rapids: Baker Academic, 2015.

———. "Register in the Greek of the New Testament: Application with Reference to Mark's Gospel." In *Rethinking Contexts, Rereading Texts: Contributions from the Social Sciences to Biblical Interpretation*, edited by M. Daniel Carroll R., 209–29. JSOTSup 299. Sheffield: Sheffield Academic, 2000.

———. “Sociolinguistics and New Testament Study.” In *Linguistic Analysis of the Greek New Testament: Studies in Tools, Methods, and Practice*, 113–31. Grand Rapids: Baker Academic, 2015.

———. “Systemic Functional Linguistics and the Greek Language: The Need for Further Modeling.” In *Modeling Biblical Language: Selected Papers from the McMaster Divinity College Linguistics Circle*, edited by Stanley E. Porter et al., 9–47. LBS 13. Leiden: Brill, 2016.

———. “Thucydides 1.22.1 and the Speeches in Acts: Is There a Thucydidean View?” *NovT* 32 (1990) 121–42.

———. *Verbal Aspect in the Greek of the New Testament with Reference to Tense and Mood*. SBG 1. New York: Peter Lang, 1989.

———. “Word Order and Clause Structure in New Testament Greek.” *Filología Neotestamentaria* 6 (1993) 177–206.

Porter, Stanley E., and Matthew Brook O’Donnell. “Conjunctions, Clines, and Levels of Discourse.” *Filología Neotestamentaria* 20 (2007) 3–14.

———. “The Greek Verbal Network Viewed from a Probabilistic Standpoint: An Exercise in Hallidayan Linguistics.” *Filología Neotestamentaria* 14 (2001) 3–41.

Porter, Stanley E., and Andrew W. Pitts. “New Testament Greek Language and Linguistics in Recent Research.” *CurBR* 6 (2008) 214–55.

Rahlfs, Alfred. *Septuagint with Logos Morphology: Rahlfs Edition*. Stuttgart: Deutsche Bibelgesellschaft, 1979. Logos Bible Software edition.

R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing, 2019.

Reed, Jeffery T. “Identifying Theme.” In *Discourse Analysis and Other Topics in Biblical Greek*, edited by Stanley E. Porter and D.A. Carson, 75–101. JSNTSup 113. Sheffield: Sheffield Academic, 1995.

Reiser, Marius. “Der Alexanderroman und das Markusevangelium.” In *Markus-Philologie*, edited by Hubert Cancik, 131–63. WUNT 33. Tübingen: Mohr Siebeck, 1984.

Reppen, Randi. “Building a Corpus: What are the Considerations?” In *The Routledge Handbook of Corpus Linguistics*, edited by Anne O’Keeffe and Michael McCarthy, 31–37. Routledge Handbooks in Applied Linguistics. London: Routledge, 2010.

Rife, J. Merle. “The Mechanics of Translation Greek.” *JBL* 52 (1933) 244–52.

Robertson, A. T. *A Grammar of the Greek New Testament in the Light of Historical Research*. 3rd edition. 1919. Reprint, Bellingham, WA: Logos, 2006.

Rüger, Hans Peter. "Die lexikalischen Aramaismen im Markus-Evangelium." In *Markus-Philologie: Historische, literaturgeschichtliche und stilistische Untersuchungen zum zweiten Evangelium*, edited by Hubert Cancik, 73–84. WUNT 33. Tübingen: Mohr Siebeck, 1984.

Runge, Steven E. *Discourse Grammar of the Greek New Testament: A Practical Guide for Teaching and Exegesis*. Lexham Bible Reference Series. Peabody, MA: Hendrickson, 2010.

Rydbeck, Lars. "What Happened to New Testament Greek Grammar after Albert Debrunner?" *NTS* 21 (1975) 424–27.

Sampson, Geoffery. *Schools of Linguistics: Competition and Evolution*. London: Hutchison, 1980.

Schmidt, Daryl Dean. *Hellenistic Greek Grammar and Noam Chomsky: Nominalizing Transformations*. SBLDS 62. Atlanta: Scholars Press, 1981.

Schmidt, Fredrick W. "Gabael (Person)." In *ABD* 2:861–62.

Schmidt, Karl-Ludwig. *Place of the Gospels in the General History of Literature*. Translated by Byron R. McCane. Columbia, SC: University of South Carolina Press, 2002.

Schneider, Edgar W. "Investigating Historical Variation." In *The Handbook of Language Variation and Change*, edited by J.K. Chambers et al., 57–81. 2nd edition. BHL. Chichester: Wiley-Blackwell, 2012.

Sinclair, John. "Introduction." In *Text, Discourse and Corpora: Theory and Analysis*, edited by Michael Hoey et al., 1–5. Studies in Corpus and Discourse. London: Continuum, 2007.

Smith, Justin Marc. *Why βίος? On the Relationship between Gospel Genre and Implied Audience*. LNTS 518. London: Bloomsbury, 2015.

Smith, Nicholas, and Elena Seoane. "Categorizing Syntactic Constructions in a Corpus." In *Research Methods in Language Variation and Change*, edited by Manfred Krug and Julia Schlüter, 212–27. Cambridge: Cambridge University Press, 2013.

Stanton, Graham. *Gospel Truth: New Light on Jesus & the Gospels*. Valley Forge, PA: Trinity Press International, 1995.

———. *Jesus of Nazareth in New Testament Preaching*. SNTSMS 27. Cambridge: Cambridge University Press, 1975.

Stadter, Philip. “Biography and History.” In *A Companion to Greco-Roman Historiography*, edited by John R. Marincola, 2:528–40. Oxford: Blackwell, 2007.

Sterling, Gregory E. “Jewish Appropriation of Hellenistic Historiography.” In *A Companion to Greco-Roman Historiography*, edited by John R. Marincola, 1:231–43. Oxford: Blackwell, 2007.

Stevens, Chris S. “Clause Structure and Transitivity: Objective Grammatical Means for Beginning Exegesis Using Philippians 2 as a Test Case.” *Conversations with the Biblical World* 35 (2015) 327–49.

———. “Clause Structuring and *Transitivity*: Advantages of OpenText.org and Its Prospects.” In *Linguistics and the Bible: Retrospects and Prospects*, edited by Stanley E. Porter, Christopher D. Land, and Francis G. H. Pang, 63–86. McMaster New Testament Studies Series 9. Eugene, OR: Wipf and Stock, 2019.

Stubbs, Michael. “On Texts, Corpora, and Models of Language.” In *Text, Discourse, and Corpora: Theory and Analysis*, ed. Michael Hoey et al., 127–61. Studies in Corpus and Discourse. London: Continuum, 2007.

“Student’s t-Test” [n.d.], 127.0.0.1:31467/library/stats/html/t.test.html.

Tan, Randall K. “Guide Through the OpenText.org Clause Annotation Project” February 3, 2006, http://opentext.org/resources/articles/a9.html.

Tardy, Christine M. “Genre Analysis.” In *Continuum Companion to Discourse Analysis*, edited by Ken Hyland and Brian Paltridge, 54–68. London: Continuum, 2011.

Thompson, Geoff. *Introducing Functional Grammar*. 2nd ed. London: Hodder, 2004.

Thomas, Margaret. *Fifty Key Thinkers on Language and Linguistics*. Routledge Key Guides. New York: Routledge, 2011.

Thumb, Albert. Die griechische Sprache im Zeitalter des Hellenismus: Beiträge zur Geschichte und Beurteilung der Koine. Strassburg: Karl J. Trübner, 1901

Tognini-Bonelli, Elena. *Corpus Linguistics at Work*. Studies in Corpus Linguistics 6. Amsterdam: John Benjamins, 2001.

———. “Theoretical Overview of the Evolution of Corpus Linguistics.” In *The Routledge Handbook of Corpus Linguistics*, edited by Anne O’Keeffe and Michael McCarthy, 14–27. Routledge Handbooks in Applied Linguistics. London: Routledge, 2010.

Tolbert, Mary Ann. *Sowing the Gospel: Mark’s World in Literary-Historical Perspective*. Minneapolis: Fortress, 1989.

Turner, C. H. “Marcan Usage: Notes, Critical and Exegetical on the Second Gospel.” In *The Language and Style of the Gospel of Mark*, edited by J.K. Elliott, 3–146. NovTSup 71. Leiden: Brill, 1993.

Turner, Nigel. *Style*. Volume 4 of *Grammar of New Testament Greek*. Edinburgh: T. & T. Clark, 1976.

———. *Syntax*. Volume 3 of *Grammar of New Testament Greek*. Edinburgh: T. & T. Clark, 1963.

Vielhauer, Philipp. *Geschichte der urchristlichen Literatur: Einleitung in das Neue Testament, die Apokryphen und die Apostolischen Väter*. De Gruyter Lehrbuch. Berlin: W. de Gruyter, 1978

Vines, Michael E. *The Problem of Markan Genre: The Gospel of Mark and the Jewish Novel*. AcBib 3. Leiden: Brill, 2002.

Walton, Steve. “What Are the Gospels? Richard Burridge’s Impact on Scholarly Understanding of the Genre of the Gospels.” *CurBR* 14 (2015) 81–93.

Watt, Jonathan M. “A Brief History of Ancient Greek with a View to the New Testament.” In *The Language of the New Testament: Context, History, Development*, edited by Stanley E. Porter and Andrew W. Pitts. LBS 6. Leiden: Brill, 2013.

Webster, Jonathan J. “An Introduction to *Continuum Companion to Systemic-Functional Linguistics*.” In *Continuum Companion to Systemic-Functional Linguistics*, edited by M. A. K. Halliday and Jonathan J. Webster, 1–11.

West, Gerald. “Judith.” In *Eerdmans Commentary on the Bible*, edited by James D. G. Dunn and John W. Rogerson, 748–57. Grand Rapids: Eerdmans, 2003.

Williamson, H. G. M. “1 Esdras.” In *Eerdmans Commentary on the Bible*, edited by James D. G. Dunn and John W. Rogerson, 851–58. Grand Rapids: Eerdmans, 2003.

Woods, Anthony, et al. *Statistics in Language Studies*. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press, 1986.

Wong, Simon S. M. *A Classification of Semantic Case-Relations in the Pauline Epistles*. SBG 9. New York: Peter Lang, 1997.

Wonneberger, R. *Syntax und Exegese. Eine generative Theorie der greichischen Syntax und ihr Beitrag zur Auslegung des Neuen Testamentes, dargestellt an 2. Korinther 5, 2f und Römer 3, 21–26*. Frankfurt am Main: Peter Lang, 1979.

Zahn, Theodor. *Geschichte des neutestamentlichen Kanons*. 2 vols. Erlangen: A. Deichert, 1889.