THE SHEPHERD OF HERMAS: A SOCIO-RHETORICAL AND
STATISTICAL-LINGUISTIC STUDY OF AUTHORSHIP AND
COMMUNITY CONCERNS.

A Dissertation by David Ian Baker.

Submitted to the Department of Religious and Theological Studies,
University of Wales, Cardiff, in Fulfilment for the Degree of Doctor of
Philosophy.

2006
ACKNOWLEDGMENTS

This Thesis owes much to the supervision of Prof. C. Trevett, Department of Religious Studies, Cardiff University. I am also indebted to Dr. T. Iles, Department of Mathematics, Cardiff University, for helpful comments concerning the statistical aspects of this thesis.
### ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANRW</td>
<td>Aufstieg und Niedergang der römischen Welt. Edited by Hildegard Temporini and Wolfgang Haase</td>
</tr>
<tr>
<td>ATR</td>
<td>Anglican Theological Review</td>
</tr>
<tr>
<td>BJRL</td>
<td>Bulletin of the John Rylands University Library of Manchester</td>
</tr>
<tr>
<td>CBQ</td>
<td>Catholic Biblical Quarterly</td>
</tr>
<tr>
<td>CH</td>
<td>Computers and the Humanities</td>
</tr>
<tr>
<td>CIG</td>
<td>Corpus Inscriptionum Graecarum</td>
</tr>
<tr>
<td>EH</td>
<td>Ecclesiastical History (Eusebius)</td>
</tr>
<tr>
<td>HTR</td>
<td>Harvard Theological Review</td>
</tr>
<tr>
<td>JBL</td>
<td>Journal of Biblical Literature</td>
</tr>
<tr>
<td>JECS</td>
<td>Journal of Early Christian Studies</td>
</tr>
<tr>
<td>JEH</td>
<td>Journal of Ecclesiastical History</td>
</tr>
<tr>
<td>JHP</td>
<td>Journal of Historical Pragmatics</td>
</tr>
<tr>
<td>JML</td>
<td>Journal of Machine Learning</td>
</tr>
<tr>
<td>JNTS</td>
<td>Journal of New Testament Studies</td>
</tr>
<tr>
<td>JTS NS</td>
<td>Journal of Theological Studies, New Series</td>
</tr>
<tr>
<td>LSJ</td>
<td>Liddell, Scott and Jones Greek-English Lexicon, (1940)</td>
</tr>
<tr>
<td>NovT</td>
<td>Novum Testamentum</td>
</tr>
<tr>
<td>NTS</td>
<td>New Testament Studies</td>
</tr>
<tr>
<td>SA</td>
<td>Sociological Analysis</td>
</tr>
<tr>
<td>TDNT</td>
<td>Theological Dictionary of the New Testament translated by Kittel-Friedrich</td>
</tr>
<tr>
<td>VC</td>
<td>Vigiliae Christianae</td>
</tr>
<tr>
<td>ZNW</td>
<td>Zeitschrift für die neuestarnentlich Wissenschaft und die Kunde der älteren Kirche</td>
</tr>
<tr>
<td>ZTK</td>
<td>Zeitschrift für Theologie und Kirche</td>
</tr>
<tr>
<td>ZWT</td>
<td>Zeitschrift für Wissenschaft Theologie</td>
</tr>
</tbody>
</table>
# CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1 Background to the Problem: Authorship and Date of The Shepherd of Hermas</td>
<td>17</td>
</tr>
<tr>
<td>2 The Quest for Authorial Style and the Rise of Statistical Linguistics</td>
<td>31</td>
</tr>
<tr>
<td>3 Model Building: Specification and Data Preliminaries</td>
<td>48</td>
</tr>
<tr>
<td>4 Unsupervised Random Forest</td>
<td>63</td>
</tr>
<tr>
<td>5 Cluster Analysis</td>
<td>74</td>
</tr>
<tr>
<td>6 Principal Components Analysis and Hierarchical Clustering</td>
<td>100</td>
</tr>
<tr>
<td>7 Optimal Variable Weighting and Minimum E-Distance Clustering</td>
<td>109</td>
</tr>
<tr>
<td>8 Finite Mixture Density</td>
<td>118</td>
</tr>
<tr>
<td>9 Working Procedure: Testing and Refining</td>
<td>125</td>
</tr>
<tr>
<td>10 Major Themes in The Shepherd: Simple Lexical Pattern</td>
<td>160</td>
</tr>
<tr>
<td>11 Conflict and Conflict Language</td>
<td>208</td>
</tr>
<tr>
<td>12 Conflict Language in The Shepherd of Hermas</td>
<td>224</td>
</tr>
<tr>
<td>13 Placing Hermas in the History of Christianity in Rome</td>
<td>235</td>
</tr>
<tr>
<td>14 Conclusions and Future Research</td>
<td>240</td>
</tr>
<tr>
<td>Appendix</td>
<td>244</td>
</tr>
<tr>
<td>Bibliography</td>
<td>258</td>
</tr>
</tbody>
</table>
INTRODUCTION

*The Shepherd of Hermas*, hereafter simply referred to as *The Shepherd*, is a long document that was highly prized in the early church. It gives an account of the visions and dreams that were experienced by the main character, Hermas. This gives the impression to the general reader that the text is of the genre of an apocalypse. While Hermas 'sees' angelic figures and the visions are explained by a spiritual guide, it lacks the visions of heaven that is central to other apocalypse literature, and also, end-of-the-world catastrophic occurrences. Consequently, *The Shepherd* cannot be considered as apocalyptic, or even pseudo-apocalyptic. The genre of *The Shepherd* will be considered in a later chapter of this thesis. A description of the narrative structure is given later in this introduction.

If the document was so highly prized in the early church, and indeed there were some who considered it to be on a par with other documents that were eventually to make up the New Testament itself, then why did it suddenly fall out of favour and be ignored by the Western church? What could have caused this dramatic and remarkable reversal of status for the text? Was it perhaps, as R.L. Fox implies, a reaction by some within the early church against a growing, or potential misuse of 'oracular service' and a blurring of distinctions between ecstatic Christianity and pagan oracles? This subject will be dealt with in chapter one of this thesis. Whatever the cause for its change of status, this was but the beginning of the problems that surround this text. Wilson, for example, concentrates on five such problems, namely, authorship, genre, canonicity, apocalyptic and the absence of the name "Jesus Christ". Modern commentators have noted more, but each new commentator has first to deal with the problem of the authorship of *The Shepherd*. This is due to the fact that the authorship and date of the text are inter-related and

---

1 Genre is a social category of communication defined by the community to which it belongs. Social convention determines the difference between genre types (eg. "history", "fiction", "comedy", "tragedy", "gospel", "apocalyptic" or "prophetic") according to the appropriate situation in which each is employed. While the boundaries between genre forms are not sharp, for example a text could be classed as 'historical fiction', there are however, certain features that would be specific to, and expect to be contained in, a text of a particular genre. There are several features that are specific to "apocalyptic" genre, including an 'other-worldly' experience, usually a vision of heaven, a hierarchy of angelic or spiritual beings, a spiritual 'guide' who explains the vision to the seer. The New Testament Book of Revelation is of typical "apocalyptic" genre. A comprehensive list of expected elements in an apocalyptic text can be found in: Collins, J.J. (ed.) *Apocalypse: The Morphology of a Genre* (Semeia 14) Scholars Press: Missoula (1979) pp.74-76

2 Fox, R.L. *Pagans And Christians in the Mediterranean World from the Second Century AD to the Conversion of Constantine*. pp.404-418

3 Wilson, J.C. *Five Problems In the Interpretation of The Shepherd of Hermas*. 

---
(given that there has been no consensus regarding the issue), each new commentator therefore needs to acknowledge their stance taken regarding the authorship and date of *The Shepherd*, in their own understanding of the work. There are two major divisions in scholarship regarding the authorship (hence date) of *The Shepherd*, and a third minority division which has been almost completely discarded. These are summarised here, but dealt with in more detail later in this thesis, as:

1) A **single author**, Hermas, who wrote the work over an extended period of time c.90 to 120 AD.

2) **Three different authors.** Author 1 who wrote *The Visions* section of *The Shepherd* sometime between 100 – 140 AD. Author 2 who wrote *Similitudes 9* around the middle of the second century. Author 3 who later added *The Mandates* and remainder of *The Similitudes*, the whole text being completed c. 160 AD.

3) **Multiple authorship.** Up to seven different authors have been suggested dating from about 70 – 130 AD. This rather complex proposition was based on an early statistical analysis of the grammar of *The Shepherd*.

This thesis develops, tests and applies a new methodology that allows these three hypotheses to be statistically evaluated and a solution of the problem of authorship (and hence date) of *The Shepherd* proposed that would aid the establishment of consensus.

The text of *The Shepherd* was preserved in a Latin translation, the Vulgate begun by Jerome c. 382 AD. Jerome also states concerning *The Shepherd*:

> Hermas, whom the apostle Paul mentions in writing to the Romans "Salute Phlegon, Hermes, Patrobas, Hermas and the brethren that are with them" is reputed to be the author of the book which is called Pastor and which is also read publicly in some churches of Greece. It is in fact a useful book and many of the ancient writers quote from it as authority, but among the Latins it is almost unknown.4

Although its popularity continued in the East (implied by Jerome above) for a while longer than in the West, by the end of the 4th century, *The Shepherd* was neglected in both the East and the West.

---

The translation is of Jerome, *De Viris Illustribus* ch10
A revival of the fortunes of the text of *The Shepherd* did not occur until 1859. Constantin von Tischendorf, whilst searching for documentary evidence for older Greek pre-cursors to the Vulgate, found a parchment text at the Monastery of Saint Catherine on Mount Sinai. This codex, later known as the Codex Sinaiticus (N), contains the texts of the Septuagint, the New Testament, the Epistle of Barnabas as well as an incomplete part of *The Shepherd*, all of which are written in Greek and probably date to the late 4th century.

The codex has been thoroughly examined for authenticity due to the discovery of a number of forged leaves of manuscript in the house of a prolific antiquities forger, Constantine Simonides. He had in his possession a few leaves of a manuscript that had been discovered in the convent of Saint Gregory on Mount Athos, which he used as a style guide for his forged documents. The genuine Mount Athos codex was later named Codex Athous (A) and probably dates to the 15th century. When arrested however, Simonides had claimed that the Codex Sinaiticus was also his work, and that he had written it himself on Mount Athos in 1839-40.

Codex Sinaiticus contains the first part of *The Shepherd*, up to Mandate 4.3.6, and Codex Athous contains almost the entire text. The partially complete text from N, and the text from A were supplemented by the Vulgate, from which Bishop Lightfoot published his translation in 1890. It was this readily available translation that finally brought *The Shepherd* into academic discussion. Therefore, the text of *The Shepherd* was reconstructed from the Greek text of N (the first quarter of *The Shepherd*, see below), the Greek text of A (which contains about 95% of *The Shepherd*) and the Latin text of the Vulgate (for the remainder of the work). There are several fragments also in existence, together with a Coptic and an Ethiopic version. Kirsopp Lake published a critical version of *The Shepherd* that contains a critical apparatus in order to compare and contrast differences between the versions. This has undergone a number of revisions since its first publication in 1913. This thesis makes use of the Greek text provided by Kirsopp Lake.

*The Shepherd* consists of three major sections represented by the headings given in the text itself, namely, **Visions** (ὤρασίς), **Mandates** (ἐντολή) and **Similitudes** (παραβολή). These major sections are further subdivided into five Visions, twelve Mandates and ten Similitudes. In order to reference individual

---

6 Lightfoot, J.B. *The Apostolic Fathers.*
7 Lake, K. *The Apostolic Fathers.* 2, pp.7-288
sections, a shorthand notation has been employed. Traditionally references were cited as, for example, Vis.2.4.3. This indicates Vision (Vis.) number 2, the fourth chapter and the third verse. Similarly Man.6.1.6 indicates Mandate 6 chapter 1 verse 6 etc. Difficulties arise when there is only one chapter, for example, Sim.1.10. This refers to Similitude 1 verse 10, because Similitude 1 consists of only one chapter. Whittaker\(^8\) introduced a new scheme of consecutive numbering where Vis.1.1.1 becomes [1], Vis.1.2.1 becomes [2] etc. Authors have tended to favour one or other of the referencing systems without consensus, and it seems to be a matter of personal choice as to which is used. Throughout the pages of this thesis I use my own scheme where V means Vision, M means Mandates, and S means Similitude. The numbers that follow are in the sequence of section, chapter, and verse. Thus V1.1.1 refers to Vision 1, chapter 1, and verse 1. Where only one chapter is present, uniformity of style requires full citation, thus, S1.1.10 would refer to Similitude 1, chapter 1 (even although it consists of only a single chapter), verse 10. This system was adopted for practical reasons of uniformity in computer programming.

The Kirsopp Lake edition of the text of The Shepherd relies mainly upon Codex Sinaiticus (N) for V1.1.1 to M4.3.6a, Codex Athous (A) for M4.3.6b to S9.29.4, and the Latin text of the Vulgate for S9.30.1 to S10.4.5. Because N is thought to be the work of at least three scribes and shows the corrections of several editors, Kirsopp Lake does not consider N to be a very reliable source. Similarly A is difficult to read and is thought to be inaccurate and not very trustworthy. Consequently, the text of The Shepherd leaves a great deal to be desired in terms of reliability. However, the less-than-perfect nature of the text presents us with interesting problems to solve. Is it possible, for example, to detect editorial insertions in the text resulting from redaction or clarification? Or, to phrase the question differently, is it possible to discover the words of the original author (or authors) in a text that may contain scribal additions, or even sections that are forgery? If a method could be devised which was sensitive enough to do this with reasonable scientific accuracy, it would prove a valuable aid in the examination of this and many other textual problems (Clementine/pseudo-Clementine, Ignatian/pseudo-Ignatian texts spring immediately to mind). Part of the aim of this thesis is to develop and evaluate such a method in order to unravel questions relating to the authorship of The Shepherd.

---

8 Whittaker, M. Der Hirt des Hermas.
Why is it necessary to undertake such a critical evaluation of *The Shepherd* at this time? It is because of the change of emphasis in histories of the early church over the past 50 years. Following Lightfoot's publication and the accessibility it gave to *The Shepherd* in academic circles, initial responses were rather dismissive. Streeter, for example, wrote concerning the work:

"Hermas is the white rabbit of the Apostolic Fathers (elderly, timid and feeble), that is why we can be certain he writes in his own name. Hermas is distinctly tedious after the first four visions."  

However, this assessment was soon to change with the advent of applications using social-scientific methodology in the study of the documents of the New Testament. In these studies, New Testament texts were regarded not only as theological statements of faith, but as witnesses to a wider social phenomenon:

"...the gathering of a community around Jesus of Nazareth conceived as Israel's Messiah and society's Saviour. This event, in turn, is comprehensible only within a larger constellation of social, economic, political, and cultural currents."  

The inevitable outcome of these methods, when applied to the context of the 1st century, was to highlight the diversity of faith and practice in the early church. It soon became customary to speak of the communities that made up the early church. Examination of Pauline, Johannine and Petrine communities soon brought interesting new insights into the beginnings of the church. Others tried to explain how these diverse communities could eventually become a 'unity' in an institutional church. Tracing the developments of such diverse communities was described as a 'trajectory', a term first suggested by Koester (borrowed from the science of ballistics when the "space race" was at its height!), which formed a simple model of change.

"Usually, the trajectory is seen in terms of a process of change from an originally radical, charismatic and creative community to a socially conformist, institutional hierarchy on its way to becoming the dominant force in the Christian church as a whole."  

---

9 Streeter, B.H. *The Primitive Church*. p.203
12 Neyrey, J.H. *An Ideology of Revolt: John’s Christology in Social-Science Perspective.*
13 Elliott, J.H. *A Home for the Homeless: A Sociological Exegesis of 1 Peter.*
However, social interactions are much more complex than this simple model will allow. The fortunes of groups within any community can ebb and flow, depending on patronage, fashion, politics and economics. As individuals within groups, or groups within communities, exercise more (or less) control, so the community itself changes and may, in some instances, fragment to form new communities. Communities may diverge, appearing to go in different directions, others may converge as they settle and mature. Some may unite for a short period in order to meet a common threat, while others may become 'closed groups' caught in an ever-tightening spiral of introspection. The practical outcome is not some smooth, well-defined trajectory from one state of community to another state of community, but a ferment of ever changing activity of re-invention and re-direction.

Nowhere is this more apparent than in the Roman church when attempting to trace a trajectory from Paul's epistle to the *Romans*, through *1 Clement* and *The Shepherd*. Taking the commonly held dates of *Romans* (c.58 AD) and *1 Clement* (c.96 AD) it would be possible to suggest that the church in Rome began to become more 'institutional' almost immediately. However, *The Shepherd* (c.70-160 AD)\(^{16}\) does not fit into this neat picture. It portrays a community organised on the earlier Pauline lines and would suggest a regression from institutional to charismatic structure! This has caused some scholars to re-date either *1 Clement*, *The Shepherd*, or both\(^{17}\). There is, however, no need to impose uniform progression upon the communities and their texts, if we accept the possibility of many different communities existing at any one time. The community from which Hermas writes is different to the community of Clement. Is it possible that Hermas' community resisted the changes that drove the Clementine community? Did they, perhaps, co-exist as equals until the Clementine community became more powerful? This raises the possibility of a very important subject in any social-scientific analysis – the role of conflict.

Conflict is a **central** facet of the socialization process and exists in all societies and all cultures. In fact, conflict is crucial to the creation of a self-identity and of community. Simmel\(^{18}\) argues that conflict in the creation of community is more crucial than even the presence of a shared cognitive substratum. Conflict

---

16 This represents the extreme range suggested by the various theories of authorship.
drives change. In written texts, the issues causing inter-community or intra-community conflict can be identified since conflict has a language of its own. Is it possible to find traces of conflict language in *The Shepherd*? This will be considered in chapter 10 of this thesis.

It is important to locate the Hermas community within the framework of early Roman Christianity and consider what were the issues that occupied their attention, and what conflicts had emerged? This thesis also addresses two related questions that follow logically from the previous one:

1. Does *The Shepherd* represent a strand of Christianity that fought against 'institutional' trends?

2. Did the Hermas community become extinct, or, did it continue as a subordinated strand that never entirely disappeared?

In the chapters that follow I will try to answer these questions and show the nature of the community and its conflict.

Before that, however, there is another issue that has to be examined because it is of vital importance in establishing the provenance of *The Shepherd*. As was noted above, there are some scholars who would question the traditional date given to *The Shepherd*. What is not so clearly appreciated is that the question of date is inextricably entwined with the problem of the authorship of the text. The conflict among scholars as to whether the work is the product of a single author or multiple authors has been a personal driving force that has led to this thesis. Most scholars are divided into two camps, those that claim single authorship (albeit over an extended period of 10-50 years) or those that claim multiple authors writing at different times. The most intriguing hypothesis was offered by Coleborne who identified six different authors. He performed a crude statistical analysis of the text (in the age before computer technology was freely available), but his theory was rejected out-of-hand as being too radical and too complex. Inspired by the pioneering work of Coleborne, this thesis applies the methods of statistical linguistics in order to shed new light on this old problem. This could now be achieved because of the work of Mealand who was able to outline a scientific, practical methodology for a statistical-linguistic study of Greek texts. A significant portion of this thesis is

devoted to the authorship problem and builds on Mealand’s work.

But although the question of the authorship of *The Shepherd* is a major concern, there are other problems that need to be addressed. These present themselves from a careful reading of the text.

**An Overview of Narrative Structure**

*The Shepherd* is a work that appears to fall neatly into three parts, namely, *Visions, Mandates* and *Similitudes*. There are five *Visions*, twelve *Mandates* and ten *Similitudes*.

*Vision 1* ("Ὀπασίς α") is concerned mainly with a vision Hermas receives of an elderly woman (γυνὴ πρεσβύτης). She sits on a great white chair (καθέδραν λευκήν μεγάλην) and reads to Hermas from a book that she has brought with her. After completing the reading she departs towards the East (πρὸς τὴν ἀνατολήν).

*Vision 2* ("Ὀπασίς β") occurs a year later with another vision of the elderly woman walking and reading from a book. She gives the book to Hermas, who is commanded to copy it and to deliver its message to God’s elect, that is, the church. The *Vision* concludes with a revelation of a young man (ἀπεκαλύφθη ὑπὸ νεανίσκος), given to Hermas while he slept. He reveals to Hermas that the elderly woman is ‘The Church’.

*Vision 3* ("Ὀπασίς γ") recounts an appearance of the elderly woman once more, and she and Hermas sit together on an ivory couch covered with linen (συμφέλιον ἐλεφάντινον...καὶ ἑπάνω λέυτιον ἐξηπλωμένον Λινοῦν καπάσιον). The purpose of this couch is to allow Hermas to sit and watch as a stone tower is built. When the elderly lady is taken away to the tower, Hermas is curious to know the meaning of the three forms in which she had appeared to him (περὶ τῶν τριῶν μορφῶν, ἐν αἷς μοι ἐνεφανίσθη). The young man supplies the information.

*Vision 4* ("Ὀπασίς δ") describes another vision of ‘The Church’ given to Hermas. This time the woman appears as young (παρωένος) and dressed in a bridal garment. This appearance occurs after Hermas has an encounter with an extremely large wild beast. The young woman explains the meaning of the beast to Hermas.
Vision 5 (Ἀποκάλυψις Ε)\(^{21}\) describes the coming of ‘The Shepherd’, whose duty is to give Hermas certain commandments and parables. The figure of ‘The Church’ does not appear in this short vision. This fact, together with the change in wording of the heading, may indicate that Vision 5 is to be regarded as having a different function in the text compared to the other Visions. It appears to serve more as an introduction to the next section, the Mandates, than one of the Visions.

Visions 1-4 seem to be a single literary unit, the woman figure of ‘The Church’ playing a prominent role in all four of them. This natural division of the Visions may have been exploited in the early 5\(^{th}\) century when they were circulated separately from the rest of The Shepherd. Indeed, The Bodmer Papyrus contains Visions 1-3 along with another vision entitled Vision of Dorotheus\(^{22}\). Carlini argued that Vision 4 was also probably attached\(^{23}\), but this cannot be proved\(^{24}\). Nevertheless, it seems that the vision material in The Shepherd was isolated from the rest of the text by an early editor specifically to bring together visionary material.

Before leaving the Visions, it should also be noted that some scholars\(^{25}\) argue for a displacement of codex pages. It is revealed in V3.10.3-5 that Hermas had seen a change in the woman’s appearance each time he had seen her.

\[\text{The first appearance is as an old woman sitting in a chair. Kirkland argues that the current Vision 1 does indeed give an account of the first appearance of the woman (V1.2.2). The second appearance of the woman is as a younger woman who...}\]

\(^{21}\) Vision 5 is headed this way in Codex Sinaiticus (4\(^{th}\) C.), but Codex Athous (15\(^{th}\) C.) heads Vision 5 as Ὀρασίς Ε, probably an editorial harmonization. However, Ἀποκάλυψις Ε is probably the best reading.


\(^{23}\) Osiel, C. The Shepherd of Hermas. p.12

\(^{24}\) The Michigan Papyrus 129 (about 250 AD), M, and some Coptic translations did not contain the Visions. This might suggest that either the Visions circulated independently of the Mandates and Similitudes which were then bound together by a later editor, or, the Visions were excised earlier and the truncated copy became that copied by M.

spoke to Hermas while she remained standing. Kirkland argues that the current Vision 3, particularly the 'Tower Vision' (V3.2.4-V3.10.6) gives an account of the second appearance of the woman, but the original beginning of this Vision has been lost. The third appearance of the woman is as young and beautiful and sitting on a couch. Kirkland argues that the current Vision 3.1.1-3.2.3, Vision 2 and Vision 4 correspond to this the third appearance. Therefore, Kirkland argues, the current Vision 2 and 'The Couch' section of Vision 3 have been misplaced. He supports this theory with a number of assumptions concerning lines of text per page, and shows that the misplaced sections correspond to whole pages, and concludes that these pages were misplaced and the page dealing with the introduction to 'The Tower' vision was lost.

Although Kirkland's theory has not found general support, being overly complicated and built on unproven assumptions, it still needs to be addressed.

Some of the objectives of the research contained in this thesis will therefore include, with respect to the Visions material, answers to the questions raised, namely:

1. Do Visions 1 to 4 show evidence of a different author to that of the Mandates and Similitudes?
2. Does Vision 5 show stylistic evidence for being an editorial seam?
3. Can any dislocation of textual narrative flow be detected in Visions 1-5 that may indicate misplaced pages of the text?

Turning now to the Mandates, it is thought that this section is considered to be a single literary unit. Each of the twelve Mandates builds on one another to form a unique manual of early Christian spirituality. Osiek26 considers their aim to be “the formation of the Christian community along the lines of Jewish wisdom paraenesis.” If this is true, then the Mandates will offer a rich source for understanding the social status of the early Roman Christian community during the period in which The Shepherd was written.

Mandate 1 sets out the foundational importance of exercising faith in, and fear of, God. Keeping this Mandate (commandment.) enables a casting off of every evil (ἀποβαλέις πάσαν πονηρίαν) in order to be clothed with every righteous virtue (ἐνδύσῃ πάσαν ἄρετὴν δικαιοσύνης) which results in a living to God (ζήσῃ τῷ θεῷ). In this Mandate the author has revealed a working principle that

26 Osiek, C. The Shepherd of Hermas. p.103
will be followed in all of the following Mandates, that of a "two ways" teaching. This form has parallels with the Didache 1-5 and the Epistle of Barnabas 18-20.

Mandate 2 deals with the virtues of simplicity (ἀπλότης) and innocence (ἀκακία). This Mandate also contains a number of verbal parallels with the Didache eg.

Πάσιν γὰρ ὁ θεὸς δίδοσθαι θέλει ἐκ τῶν ἰδίων δωρημάτων. (M2.1.4)
Πάσι γὰρ θέλει δίδοσθαι ὁ πατήρ ἐκ τῶν ἰδίων χαρισμάτων (Did 1.5)

This has raised questions of the possible literary dependence of the Didache on The Shepherd, or of both on a common source.

Mandate 3 deals with the subject of truthfulness (ἀληθεία).

Mandate 4 seems to have a dislocation of the text. M4.1.1-11 deals with questions about adultery, while M4.1.4-4 deals with the question of remarriage after the death of one partner. There is an obvious continuation of theme between these two sections. However, between these two sections there is M4.2.1-M4.3.7 which deals with concerns about repentance. Osiek claims, "It has been suggested that the two middle chapters were originally a continuation of Vision 5 before its last two verse"27, but she does not cite who first suggested this.

Mandate 5 describes two modes of living based on indwelling of the Holy Spirit, who brings cheerfulness (ὕλαροτη) and well-being (εὐθυνία), and that of the indwelling of an evil spirit who causes ill temper (ὀξυχολία).

Mandate 6 continues the theme in Mandate 1 with modification of the two spirits theme of Mandate 5. M6.2.1-4 describes two angels that dwell with humankind, one the 'Angel of Righteousness', the other the 'Angel of Wickedness'.

Mandate 7 continues the theme of 'fear of God', first introduced in Mandate 1, expanding the discussion to show two types of fear.

Mandate 8 encourages self-control/restraint (ἐγκράτεια). Again, the discussion proceeds on the basis of two types of restraint.

27 Osiek, C. The Shepherd of Hermas. p.109
Mandate 9 deals with the subject of double-mindedness (διψυχία). Osiek believes that this Mandate belongs with the fifth and tenth in terms of unity of theme.

Mandate 10 describes the state of grief (λύπη) as being the result of the other vices mentioned in Mandates 5 and 9. Therefore, Hermas encourages a state of joyfulness (λαροτησ), which is more acceptable to God. We see that a two-ways theme is again developed in this Mandate also.

Mandate 11 continues the two-ways theme in relation to prophets and prophecy, by comparing ‘true’ and ‘false’ prophets.

Mandate 12 deals with good and evil desires (ἐπιθυμία) before concluding with M12.3.2. However, Hessens continues discussions with ‘The Shepherd’ until the Mandates close at M12.6.5. Therefore this section, M12.3.3 to M12.6.5, may be an editorial insertion when the Mandates were separated from the Similitudes.

This overview over the Mandates has raised some additional questions that will need to be addressed, namely:

4. Is there evidence of dislocation in the text, particularly M4.2.1 - M4.3.7?
5. Is M12.3.3 to M12.6.5 written by a different author/editor?

The final section of The Shepherd is a collection of Similitudes (parables). There is no indication that the Similitudes had a separate introduction, rather, they appear to continue the themes developed in the Mandates. V5.1.5 is the only introductory link.

Similitude 1 is a parable of two cities, a theme dealing with the conflict of interest the Christian has between earthly and heavenly citizenship. It is a theme that is also found in Hebrews 13:14, but in The Shepherd it is linked to the idea of obtaining wealth. Hermas exhorts his readers to use any wealth they have for the benefit of others:

εἰς τούτο γάρ ἐπλούτισεν ὑμᾶς ὁ δεσπότης, Ἰνα ταῦτας τὰς διακοινίας τελέσητε αὐτῷ. (S1.1.9)

Similitude 2 describes the parable of two vines. Although this parable has been used as an indicator of Roman provenance of The Shepherd, its purpose in the text is to remind the rich members of the community to care for and support the poorer members.

28 Dibelius, M. Der Hirt des Hermas. p.546
Similitude 3 and Similitude 4 are a parable of the two seasons, winter and summer, and their effects upon the foliage of trees. Hermas likens this world to a winter season, whereas, the world to come is likened to the summer for the righteous, when, (τῶν δικαίων οἱ καρποὶ φανεροὶ).

Similitude 5 seems to be a strange mixture of themes and ideas. It begins with questions related to fasting (S5.1); a parable about a vineyard follows (S5.2), which is a parable about 'true fasting'. This becomes apparent in S5.3 where the explanation of the parable is given. True fasting and care for the poor must go hand in hand. Difficulty arises in the fact that the parable itself mentions the vineyard owner's son. However, in the explanation of the parable in S5.3, the son is not mentioned. The son is mentioned in the allegorical interpretation in S5.4 and the remainder of the Similitude deals with the son-slave relation with respect to the Son of God. This raises the question as to whether a vineyard parable relating to fasting and a vineyard parable relating to the role of the son have been fused together in this Similitude, either by Hermas or a later editor.

Similitude 6 and Similitude 7 are a parable of two shepherds, 'The Shepherd of Luxury' and 'The Shepherd of Punishment', although it is the sheep that are the focus of the parable. Osiek suggests that these two Similitudes offer an explanation of suffering.

Similitude 8 is a parable about the cosmic willow tree. The different kinds of branches parallel the different kinds of stones in the tower vision.

Similitude 9 is the longest similitude. It opens with Hermas being transported to Arcadia where he sees twelve mountains (S9.1) but the explanation of the mountains is not given until S9.17 ff. Sandwiched between this vision and its explanation is another vision. In this vision Hermas is shown again a vision of the building of the tower. Because of this second reporting of the tower building, and because of the length of Similitude 9, it is interesting to speculate whether this Similitude once circulated independently of the other Similitudes. If so, it might have been added to the other Similitudes by a later redactor.

The Similitudes therefore raise further questions that need to be answered. In particular:

6. Do the Similitudes display the same characteristics of authorship as the Mandates?
7. Does Similitude 5 show evidence of a different author or the presence of editorial links?

8. Does Similitude 9 show evidence of a different author or the presence of editorial links?

9. Does the first section of Similitude 5 have stronger links with the Mandates than with the Similitudes?

Therefore, including the question of authorship, there are 10 questions that need to be considered at various points in this thesis.

This thesis is divided into two parts. The first part deals with the problem of the authorship and date of *The Shepherd* and the method used to identify authors must be scientifically established. The solution of this problem is essential for any understanding of the social location of Hermas. If the text is composed by several authors, presumably over an extended period of time, then it will be a complicated task of un-picking the text in order to place Hermas in the history of Christianity in Rome. If single authorship is found to be the conclusion of the analysis, then it will be an easier task to place Hermas in a social location. Therefore, the second part of the thesis deals with issues of social location and community, based upon the findings of the first part of the thesis.

The major emphasis, then, is on part one of the thesis and the development of a method that will enable detection of single or multiple authors. This is based on multivariate statistical methods, and it must always be remembered that statistics deals with *probabilities*. Therefore, statistical analysis cannot be used in isolation from other forms of critical textual analysis. However, part one will develop a statistical methodology that uses a different approach to the one that is normally used by those performing linguistic statistics.

There are particular presentational difficulties associated with this section because, this is not intended as a mathematics thesis, but one for Humanities studies. It falls, then, between two camps. In order to enable those without an interest in mathematics to understand the ideas of each section, and to participate in its conclusions, I have prefaced each mathematical section with a simplified non-mathematical introduction. It is hoped that this approach will not offend either mathematical or non-mathematical readers.

This thesis can be divided into two parts. Part one deals with the question of
authorship and the application of multivariate statistics. Part two deals with the Hermas’ community, its status and conflict, as implied in the text of The Shepherd. These themes are extracted from the text by application of grammatical-linguistic methods applied to the text. The thesis outline is therefore as follows:

**PART ONE (Authorship)**

Chapter 1 outlines how the problems relating to the authorship and date of The Shepherd have arisen. It is proposed that dispute about authorship is mainly due to the credence that some scholars give to the Muratorian Fragment. The debate concerning the Muratorian Fragment is bound up with the ecclesiastical politics of the first three centuries and also with the ecclesiology of the 19th and 20th centuries. It is, in fact, a false trail that has led to erroneous conclusions, but has raised important issues concerning the reliability of internal textual evidence and its relationship to the authorship question.

Chapter 2 explains the historical development of text-statistical methodology including the failures and the advances of technique.

Chapters 3 to 9 deal with the advancement of a methodology to deal with the problem of the statistical measure of authorship. They seek to show the unsuitability of current accepted practice, and propose a new group of techniques that are better suited for the purpose. The methodology is tested for robustness and sensitivity in relation to the tasks concerned.

**PART TWO (Community)**

Chapter 10 widens the grammatical appreciation of the text of The Shepherd by examining the themes that are developed through the text. This makes use of lexical repetition and the theories of Michael Hoey. However, lexical repetition has been applied mainly to English language texts, while in this thesis it will be shown that it applies equally well to Greek language texts. This chapter will also help to answer some of the questions posed from the careful reading of the text (see above).

Chapter 11 deals with a grammatical approach to the study of conflict and conflict language. It introduces current theories of conflict and a local grammar of conflict language. These studies were originally developed for modern English studies and based on corpus studies; consequently the feasibility of application of the method to 1st century Greek texts must be critically examined.
Chapter 12 applies the local grammar of conflict to a study of *The Shepherd* in order to assess the main concerns of the Hermas community in Rome.

Chapter 13 brings together the information gleaned from the text about Hermas and his community (chapters 10 – 12 of this thesis). A synthesis of this information with that of an understanding of Christianity in Rome will place Hermas in his social context. Hermas’ community is shown as one that is in a process of decline. The chapter suggests why this situation had arisen and what the outcome was for the Hermas community.

Chapter 14. Conclusions and points relevant for future research.
CHAPTER 1

BACKGROUND TO THE PROBLEM:

AUTHORSHIP AND DATE OF THE SHEPHERD OF HERMAS

Introduction

Although The Shepherd of Hermas had been a prized text in some sections of the early church during the first three centuries, it suddenly lost its popularity and was marginalized by the fifth century. The purpose of this chapter is to discover why it underwent this reversal of fortune.

Wilson discusses the fortunes of The Shepherd over a number of centuries concerning its acceptance and rejection as scripture. He used evidence from Clement of Alexandria, Irenaeus, Origen, Tertullian, Eusebius of Caesarea, The Muratorian Fragment, Codex Sinaiticus and a number of canonical lists. Despite this scholarly review, he concludes that The Shepherd was probably excluded from the canon of the New Testament because of its length.

I believe that my third reason, its far greater length than any canonical New Testament document may be the reason for its evolving out of the canon, even though no church father ever gives this as a reason. (bold mine)

Is the length of the text likely to be the major factor in its exclusion from the canon and its decline in popularity? It may be a reason, but probably just one of many reasons for the decline in popularity of The Shepherd; the reason must be more complex than that alone.

Wilson’s Evidence Reconsidered.

1. Clement of Alexandria (c.150 – c.215 AD)

Clement quotes freely from The Shepherd and considers it to be scripture, but he also quotes from apocryphal Acts and the Apocalypse of Peter, indicating that there is no thought of a closed canon for Clement. Wilson regards Clement as a typical example of the first of three stages toward the process of deriving the canon.

29 Wilson, J.C. Five problems in the Interpretation of the Shepherd of Hermas. pp.51-72
30 ibid. p.70
This first stage he describes as the inclusive stage, characterised by the emergence of the idea of inspired writing and openness to the possibility of a wide range of inspired Christian writing. Clement did not leave us a list of texts that he thought were canonical, or a list which were not canonical, rather, he was open to deriving Christian meaning from all types of texts, because Clement was one who developed and used an allegorical interpretation of Christian writings. *The Shepherd* was ideally suited to his method of interpretation; consequently it is not surprising that he quotes it so often in his own works. This does not mean that Clement should be classed in the first stage of a developing process toward a closed canon. What it does tell us is that *The Shepherd* found a ready audience in Egypt, where, according to Osiek31, “more fragments of it have been found thus far than any other non-canonical text”.

2. Irenaeus of Lyons (c.142 – c.200 AD)

Irenaeus was noted for his works challenging the Gnostics, whose doctrines had achieved a wide acceptance in the second century. According to Eusebius, Irenaeus wrote a list of texts he considered non-canonical, and giving reasons why. As far as his thoughts about *The Shepherd* are concerned, Irenaeus does quote from M1.1.1. where he says (according to Eusebius):

```
άλλα καὶ ἀποδέχεται τὴν τοῦ Ποιμένος γραφὴν λέγων, καλῶς οὖν ἡ γραφή ἡ λέγουσα πρῶτον πάντων πίστευσον ὃτι εἰς ἐστιν ὁ θεὸς ὁ τὰ πάντα κτίσας καὶ καταρτίσας
```

This is the only quote Irenaeus takes from *The Shepherd*, but it is important because it suggests that he regarded the text as scripture. Wilson then gives his conclusion that, “Though he (Irenaeus) may consider it scripture, it does not seem to him to be very important scripture.”33 This rather dramatic conclusion is necessary because Wilson wants to show that Irenaeus falls into a second stage of the process of deriving the canon of the New Testament. According to Wilson’s scheme, stage two is the time when the idea of fixing and closing a canon of scripture has not yet taken hold, but there is an increasing awareness of heresy and an emerging orthodoxy.

Wilson, of course, cannot possibly know something as subjective as the level

31 Osiek, C. *The Shepherd of Hermas*. p.6
32 Irenaeus in Lake, K. (tr.) *Eusebius Ecclesiastical History (Books I to V)*. p.456
33 Wilson, J.C. *Five problems in the Interpretation of the Shepherd of Hermas*. p.59
of importance with which Irenaeus regarded *The Shepherd*. All we know for certain is that Irenaeus had read at least part of the text.

3. Origen (c.185 – c.254 AD)

Origen, like Clement, is an advocate of the allegorical interpretation school of thought. Like Irenaeus he also quotes Mandate 1.1, but the most revealing comment about *The Shepherd* is to be found in *De Principiis*:

> For as man consists of body, and soul, and spirit, so in the same way does Scripture, which has been arranged to be given by God for the salvation of men. And therefore we deduce this also from a book which is despised by some—*The Shepherd*—in respect of the command given to Hermas to write two books, and after so doing to announce to the presbyters of the Church what he had learned from the Spirit. The words are as follows: "You will write two books, and give one to Clement, and one to Grapte. And Grapte shall admonish the widows and the orphans, and Clement will send to the cities abroad, while you will announce to the presbyters of the Church.

Although it is not clear whether Origen regards *The Shepherd* as scripture, it is clear that he has heard that some do not regard it very highly. He gives no indication of who despises the text, or why they despise it.

Sometime between Clement of Alexandria and Origen there were those, presumably in the church, who had begun to question the value of *The Shepherd*. While it may be conceded that the word ‘despise’ may be a rhetorical exaggeration, surely it would not be a word, or idea, that would be used by those who felt that *The Shepherd* was too long! Rather, it is a word that would better fit those who disliked the content of *The Shepherd*. We can therefore suggest that between c.200 –250 AD there were some who argued against the authority of *The Shepherd*.

Origen continued to hold *The Shepherd* in high regard because he believed that the Hermas mentioned by Paul in Romans 16:14 was the Hermas of *The Shepherd*. Wilson notes that Origen in his commentary on Romans writes, “Hence I think that this Hermas is the writer of the little book which is called the Shepherd, and I think it is divinely inspired.”

4. Tertullian (c.160 – c.225 AD)

In his earlier works, Tertullian quotes from *The Shepherd* and regards it as

---

35 Quoted in Wilson, J.C. *Five problems in the Interpretation of the Shepherd of Hermas.* p.63
When he became a Montanist, his attitude to The Shepherd also changed. He disliked what he saw to be its moral and spiritual laxity, and refers to it as follows:

> And, of course, the Epistle of Barnabas is more generally received among the Churches than that apocryphal "Shepherd" of adulterers...  

He dislikes the “shepherd of adulterers” because it can countenance and advocate repentance even after divorce for adultery. More importantly, he tells us that The Shepherd was not well received in some churches.

Wilson places Tertullian at the end of stage two of his historical process of deriving the canon of the New Testament. However, it helps us to narrow the period during which the authority of The Shepherd was under question to the period c.200 – c.225 AD.

5. Eusebius of Caesarea (c.260 – c.340 AD)

By the time of Eusebius the status of The Shepherd had completely changed. He writes:

> ἐπεὶ δ’ ὁ αὐτὸς ἀπόστολος ἐν ταῖς ἑπὶ τέλει προσφησει τὴς πρὸς Ἡρωμαίους μνήμην πεποίηται μετὰ τῶν ἄλλων καὶ Ἐρμά, οὐ φασιν ὑπάρχειν τὸ τοῦ Ποιμένος βιβλίον, ἱστέρου ὡς καὶ τούτῳ πρὸς μὲν τινὸς ἀντιλέκται, δι’ οὓς οὐκ ἄν ἐν ὑμνολογομένοις τεθεῖν, ὅσ’ ἐτέρων δὲ ἀναγκαίωται οὓς μᾶλιστα δεὶ στοιχείσεως εἰσαγωγικῆς, κέκριται. ο.pen ήδη καὶ ἐν ἐκκλησίαις ἴσμεν αὐτὸ δεδημοσιευμένον, καὶ τῶν παλαίτατων δὲ συγγραφέων κεχρημένοις τινάς αὐτῷ κατείληφα.

It is now obvious that The Shepherd is no longer considered authoritative so as to be placed with the accepted books, but it is still popular and used for catechesis in some churches.

Wilson places Eusebius in his third stage of canon development, the ‘exclusive stage’ when books that contain discrepancies of doctrine are excluded from the canon.

37 See for example the section in the Mandates, M4.1.6-8
38 Lake, K. Eusebius Ecclesiastical History. 3.3.6 p.193
If Wilson’s scheme is correct, then *The Shepherd* entered stage 2 (the transitional stage) very early in its history. Moreover, it was despised by some, which suggests that its content, not its length, caused offence to some. If it might be established what section of the church opposed its message and eventually undermined its authority and what the message was that was unacceptable to some sections of the church, then it might be possible to identify the underlying reasons for its rejection. Similarly, we need to establish the significance to the period c.200 – c.225 AD when the despisers were actively engaged against the authority of *The Shepherd*.

Next we turn to a crucial piece of evidence that will help us to answer these questions, the Muratorian Fragment.

**The Muratorian Fragment**

(A translation of the Fragment can be found in Appendix 1.1)

The Muratorian Fragment\(^\text{39}\) was first discovered in 1740, and published by the eminent textual critic S.P. Tregelles in 1867, who believed it to date to the late second century. It gives a list of books of the New Testament considered to be canonical by the author of that Fragment (usually referred to as the Fragmentist), and a list of books that the author argues are not canonical. Since The Fragment appears to make a direct temporal reference to *The Shepherd* (see below) it was used to assign a date of between 150-200AD to *The Shepherd*. However, it soon became apparent that not all scholars found this dating of *The Shepherd* to be acceptable and in accord with the evidence found in the text itself. This, in turn, raised serious questions about the date of The Fragment.

The Fragment is widely considered to be of Roman origin, but there have been a number of writers\(^\text{40}\) have argued against this, preferring either Syria or Palestine as its area of origin. Those who suggest a provenance other than Roman point to the fact that 1 Peter is not mentioned in the list of approved books. This does seem a surprising omission, but Westcott overcame this criticism by suggesting that since the

\(^{39}\) Metzger, B. *The Canon of the New Testament*. pp. 191-201 provides a good overview of this document.

Fragmentist could not possibly have missed out reference to 1 Peter (or even James and Hebrews) then the Fragment must have been mutilated in some way. Many likely explanations for the disruption in the text have been put forward, but of course; the content of any missing text is only speculation.

One other peculiarity of the Fragment that requires immediate comment is the quality of its Latin grammar. Although the Fragment is extant only in Latin, the quality of its Latin is so poor that some consider it to be a bad translation of an original Greek text. This has implications with regard to its dating. Hill deals with the Latin word 'catafygum' (line 84 of the Fragment) and Hahneman argues that this is a translation of the Greek Κατάφυγας which was a name given to Montanists in the 4th century, therefore he proposed that this was a more likely date for the Fragment, rather than the commonly accepted date of the late 2nd century. Hill and Henne both suggest that this could be the work of the Latin translator, and if that is correct, then the Latin translation could be of the 4th century, but that does not mean the original Greek version is also 4th century.

There appears to be a temporal link in the Fragment itself:

_We receive only the apocalypses of John and Peter, though some of us are not willing that the latter be read in church. But Hermas wrote the Shepherd very recently, in our times, in the city of Rome, while bishop Pius, his brother, was occupying the chair of the church of the city of Rome. And therefore it ought indeed to be read; but it cannot be read publicly to the people in church either among the Prophets, whose number is complete, or among the Apostles, for it is after [their] time._

The Fragment seems to suggest that Pius and Hermas (elsewhere in the document Marcion and Montanus also) are contemporaneous with the Fragmentist. This would argue a late 2nd century date for The Fragment. Debate has surrounded the phrase "quite lately in our times" (nuperrim e(t) temporibus nostris), especially the usage of the term 'nuperrim'. Dates of between 150 and 200 have been suggested as fulfilling the criteria of 'in our times'. Zahn preferred a date toward the end of this range because he believed the Fragment was rejecting Montanist writings.

41 Lightfoot, J.B. _Apostolic Fathers._ 1, pp.261-266
43 Hahneman, G.M. _The Muratorian Fragment and the Development of the Canon._ pp.25-26
44 Henne, P. "La datation" pp.63-64 – Quoted in Hill.
Along with a couple of Marcionite epistles and some Gnostic writings the author of the fragment implies that the readers of The Fragment knew The Shepherd, and that it was read in some churches. The whole thrust of Hahneman's argument is that a list of canonical writings could not have occurred before the 4th century. However, the formation of a closed canon assumes that some sort of process of development had previously taken place. It is possible, then, that lists of acceptable books had been written before a definitive acceptance of a final canon was proposed (as suggested by Wilson's three stage scheme). Not only could such lists have been in use, but also different churches may have adopted slightly different lists, thus precipitating the need for an acceptable agreement of a closed canon. More importantly, the Fragmentist shows great antagonism with the content of The Shepherd, as displayed in the remark, "it cannot to the end of time be publicly read in the Church". However, private reading is conceded. If the author wanted the complete rejection of The Shepherd from the church (as with the works of Marcion), he finds that it cannot be easily achieved because it still remains popular in some Christian circles. The concession to private reading is therefore given. The author also excludes The Shepherd from the canon of the New Testament because it cannot be placed among the prophets or the Apostles. Ferguson rightly concludes that it is hard to imagine why a 4th century author would invent a 2nd century persona to discredit The Shepherd.

Although the arguments for a 4th century dating of The Muratorian Fragment have been proposed, they are not very convincing and a 2nd century date seems more probable. The provenance is probably, though not certainly, Roman. We can now assess how this bears upon the problem of the date and authorship of The Shepherd.

The Problem of Ecclesiastical Politics

Taken at face value, the Muratorian Fragment ascribes single authorship to The Shepherd and places it in the time frame c.140 AD to c.150 AD, the approximate dates for Pius' episcopate. If the Muratorian Fragment is a late 2nd century document of c.190 AD, it implies that The Shepherd was not being read, and was being actively restricted in certain sections of the church, as early as the c.200 AD, which agrees with the time already suggested in the previous section (c.200 – c.225

AD). The Fragmentist also implies it was still being read in other sections of the church and was still highly prized. We have already noted that this was the case when we examined the evidence from Clement of Alexandria, Irenaeus, Origen and Eusebius. Athanasius of Alexandria (c.296 – c.373 AD) in his 39th Festal Letter writes:

"But for the sake of greater accuracy I add, being constrained to write, that there are also other books besides these, which have not indeed been put in the canon, but have been appointed by the Fathers as reading-matter for those who have just come forward and which to be instructed in the doctrine of piety: the Wisdom of Solomon, the Wisdom of Sirach, Esther, Judith, Tobias, the so-called Teaching [Didache] of the Apostles, and the Shepherd. And although, beloved, the former are in the canon and the latter serve as reading matter, yet mention is nowhere made of the apocrypha; rather they are a fabrication of the heretics, who write them down when it pleases them and generously assign to them an early date of composition in order that they may be able to draw upon them as supposedly ancient writings and have in them occasion to deceive the guileless." 47

It would appear that the churches in North Africa continued to value the work even into the 4th century, while it was being devalued elsewhere.

This data begins to make sense when it is seen as a result of an interaction with Montanism, initially in Asia, but also in Rome. If we take Epiphanius’ dates as correct, Montanism began to attract attention around 156/157 AD and it spread rapidly through Asia. At first it was not regarded as heretical, but seemed to contain elements that could be traced back to Christian beginnings (Acts 2) and the fulfilment of an older promise (Joel 2:28-29). Harnack suggests the appeal of the teaching was its enthusiasm and primitive fervour in the face of a growing institutionalised church. If this is so, Montanism was on a collision course with some of the Christian communities in Rome. Eusebius48 may allude to this in a short comment that he makes:

τῶν δ’ ἁμφί τῶν Μοντανῶν καὶ Ἀλκιβιάδην καὶ Θεόδοτον περὶ τὴν Φρυγίαν ... ἡς ἐν δεσμῷ ἔτι ὑπάρχοντες τοῖς... ἔλευθέρω τῷ τότε Ῥωμαίων ἐπισκόπῳ, τῆς τῶν ἐκκλησιῶν εἰρήνης ἐνεκα πρεσβεύοντες.

48 Lake,K. Eusebius Ecclesiastical History. 5.3.4. p.443
If "the most plausible interpretation" is that Eleutherus’ attitude to Montanists in Rome was one of opposition, then divisions were already appearing openly between 174 and 189 AD. By 200 AD Gaius was able to write an anti-Montanist work in Rome.

What bearing does this have on *The Shepherd*? Within the Montanist movement women played a prominent role. Visions and ecstatic utterances were seen as a means of communication flowing from the spiritual realm to the mortal world. Epiphanius, preserving an early anti-Montanist source, gives us some idea of the sort of communication that was received. Although these were mostly in the form of ecstatic utterances, Epiphanius also gives evidence of an alleged vision, which has relevance in its similarity to Hermas' vision of the Woman Church:

*Phrygians, also called Montanists and Tascodrugians. They accept the Old and the New Testaments but, by boasting of a Montanus and a Priscilla, introduce other prophets after the [canonical] prophets.*

*Pepuzians, also called Quintillianists, with whom Artotyrites are associated. They derive from the Phrygians but teach different doctrines. They venerate Pepuza, a deserted city between Galatia, Cappadocia and Phrygia, and regard this as Jerusalem. (There is another Pepuza as well.) And they allow women to rule and to act as priests... And they tell the story that Christ was revealed in female form to Quintilla, or Priscilla, there in Pepuza.*  

(Attributed to Priscilla or Quintilla).

Hermas’ visions included one of Rhoda, his former owner, the Woman Church, and seven maidens, consequently it is possible that anti-Montanist zeal would cause *The Shepherd*, containing these prominent women figures, to be regarded with suspicion. For the anti-Montanist party the argument was crucial and was bitterly fought. Why?

Trevett is correct in her assessment that the main debate between Montanism and the catholic church was really about authority and power. Who had the right to act as the mouthpiece of the Divine? The catholic church answered this question by resorting to a uniformity of practice and revelation, and for them written records were necessary for uniformity of practice, discipline and control. They regarded

---

49 Lampe, P. *From Paul to Valentinus.* p.394, note 25
50 Epiphanius *Haer.* 48,49 in Williams, F. (trs.) *The Panarion of Epiphanius.* In the refutation of 80 heresies, Epiphanius gives examples of Montanist uttering.
51 See Trevett, C. *Montanism.* pp.167-171 for a discussion of whose prophecy this is.
52 Trevett, C. *Montanism.* pp.140ff
Montanism as representing uncontrollable and unchecked practice, even though the Montanists had their own writings, and could appeal to the same writings as the catholics. However, authoritative writings could be added to their collections without limit. For the Montanists then, scripture was a living and growing experience of the spirit's work in the church. The catholic sections of the church sought to counter the many private individual prophecies and visions by considering a closed canon of acceptable writings. Not everyone agrees with this assessment, Paulsen\textsuperscript{53}, for example, believes Montanism's effect on canon creation to be overestimated. The contest was an ecclesiastical political power struggle. Therefore, any writing that might be seen as containing text that could be interpreted as favourable to a Montanist or ecstatic viewpoint would be treated with suspicion. The Shepherd with its vision of the woman church (compare the description of Priscilla's vision of Christ with Hermas' vision of the woman church in Vision 4.2.1) and teaching conveyed by an angelic figure seemed too close to a Montanist position to be allowed a public reading in the church. It is interesting to note that the Fragmentist accepts Revelation, but comments that there were some who would not allow it a public reading.

But Revelation was regarded as less doubtful than The Shepherd, even if many of the elements in the narratives were similar eg. vision guides, angels, beasts, tribulation, books/letters to be written and circulated amongst the churches. Revelation had two strong supports, first many held its apostolic authorship (unlike The Shepherd) and second the major vision figure of Christ was as 'a son of man' (unlike the Visions in The Shepherd).

But although there followed a distancing of the catholic church from The Shepherd, it could not be an outright rejection because it had already gained a standing and, as we have seen, a few eminent supporters in the church.

Eusebius gives us a clue to the campaign of the anti-Montanist parties:

\[\text{τὴν δὲ καθόλου καὶ πᾶσιν τὴν ὑπὸ τῶν οὐρανῶν ἐκκλησίαν βλασφημεῖν διδάσκοντος τοῦ ἀποκαλυθησμένου πνεύματος, ὅτι μήτε τιμῆν μήτε πάραδον εἰς αὐτὴν τὸ θεουσπροφητικὸν ἐλάμβανε πνεῦμα, τῶν γὰρ κατὰ τὴν Ἀσίαν πιστῶν πολλάκις καὶ πολλαχῇ τῆς Ἀσίας εἰς τούτο συνελθὼν καὶ τοὺς προσφάτους λόγους}\]

\textsuperscript{53} Paulsen, H. “Die Bedeutung des Montanismus für die Herausbildung des kanons.” \textit{Vigiliae Christianae} (1978) 32, pp.19-52
Montanism seems therefore to have been condemned by the Asian churches and eventually by Zephyrinius of Rome (d. 217) who was also persuaded to ratify their decision.

It is, therefore, entirely possible that the Muratorian Fragment represents the formal guidelines of the anti-Montanist party in Rome c.190 AD. It is not necessary to go along with Sundberg and re-date The Fragment to the 4th century. Sundberg proposes this date on the basis that all the lists of canonical books up to the time of Eusebius (early 4th century) accept The Shepherd as canonical, and that all the lists after that time reject it. Therefore, he argues, The Muratorian Fragment which rejects The Shepherd of Hermas, must be of 4th century date. According to Sundberg's logic it is only when everyone else 'catches-up' and agrees with the statements expressed in The Fragment concerning The Shepherd that The Fragment can then be said to exist! Sundberg has been fooled into the trap of expecting uniform progression of the church as if it were a uniform single entity. In the confrontation with Montanism we see a group within the church taking a hard-line anti-Montanist position that moves quickly to attack the perceived threat and to strengthen its own defences. The churches in North Africa do not see Montanism as so much of a threat and they have no suspicion of The Shepherd, a work that they valued highly.

Bearing in mind the tension, it now becomes clear why the author of The Fragment ascribes authorship of The Shepherd to a relatively recent author, namely Hermas the brother of Pius. If there is no apostolic connection there can be no good reason for giving it canonical status and its authority can be questioned. This prevents groups in the church who support ecstatic experiences using The Shepherd as an authoritative source to support their position. On the other hand, Origen, still wants to support The Shepherd because of the suggested apostolic connection with Paul which gives the work authoritative status.

54 Lake, K. Eusebius Ecclesiastical History. V. 16.10. p. 477
55 Trevett, C. Montanism. pp. 55-65
57 Indeed, Tertullian seems to have remained catholic while also being Montanist. See Trevett, C. Montanism. pp. 66-76
Therefore, the interaction of Montanism with the catholic churches in Rome can be seen to account for the available evidence we have for the decline in fortune of *The Shepherd*. It explains why it was despised in some circles because of the prominence of visions of women, making it appear too close to a Montanist position. It explains why the decline in its popularity occurred at the period suggested by the writings of the Church Fathers. It also explains why its use in some geographical sections of the church continued for longer than others.

Returning to the question of authorship, there are only two contenders, and consequently, only two periods for its production:

1. Hermas the brother of Pius, and therefore a date for the writing of the document in the range 130 - 150 AD.

2. Hermas, the Roman Christian leader already active in Rome before Paul's visit to the city. A date somewhere in the range 70 - 105 AD is possible.

It can be seen that both of these options imply single authorship. Multiple authorship becomes a necessary hypothesis only when trying to combine these two positions, as we shall now see.

**The Problem Of Synthesis**

The conflict in these two proposals about authorship is the cause of the uncertainty expressed by modern commentators. If equal weight is given to both proposals it becomes a task of synthesis.

In 1963, Giet offered his 3-author hypothesis. It was his attempt to give equal weight to both proposals and he suggested the following scheme:

**Author 1** - the Hermas of the 1st century who wrote Visions 1 to 4.

**Author 2** - the Hermas who was brother of Pius and wrote Similitude 9.

58 Although the consensus is moving toward single authorship, which is supported by Joly and Barnard for example, there are still a few scholars attracted to multiple author theories. Osiek originally supported a three author hypothesis, but now favours a single author writing over an extended period of time.
Author 3 - a pseudo Hermas who wrote Vision 5, the Mandates, Similitudes 1 to 8 and similitude 10. This author completed the work by about 161 AD.

Giet\(^\text{59}\) initially convinced many commentators not only because his theory offered the advantage of satisfying all of the dating evidence, but also, by his careful division of the text he managed to resolve some of the internal difficulties presented by the text. We will examine the internal evidence shortly.

However, Giet failed to convince many others who felt there was no evidence of multiple authorship. Reiling\(^\text{60}\) argued for a single author who "wrote the various parts of the book at different times", although he gives no indication of date range. Joly, Jeffers, Barnard, Wilson and Henne all argue for a single author. Jeffers and Barnard suggest the work began with Visions 1 to 4 that were written at the close of the 1st century and the remainder of the work was then completed about 135 AD.

Those who advocate a single authorship hypothesis do so with various hedges, such as, that multiple sources were used, or multiple redactions, or an extended period of writing. Most commentators, holding whatever hypothesis, are generally agreed that *Visions 1-4* form a unity and that *Vision 5* is probably an introduction to the *Mandates*. These arguments are based upon the commentators’ understanding of the content of the text.

In his work of 1969, Coleborne\(^\text{61}\) approached the task differently. He based his work on the grammatical construction of the text, and argued that different authors would use different grammatical styles. He believed that grammar could be used as a measure of authorship. After grammatically tagging every word of the text he analysed the frequencies of grammatical construction as they occur through the text. Not only did he measure the frequency of adjectives, but he also noted those occasions when the author placed the adjective before the noun, and those cases where the adjective was placed after the noun. He arrived at the conclusion that there were six different authors involved in the writing of *The Shepherd*. These he identified as:

**Author 1** - who wrote the Mandates 1 to 12.3.3. This was the earliest section and could be dated about 70 AD.

\(^{59}\) Giet, S. *Hermas et les pasteurs*

\(^{60}\) Reiling, J. *Hermas and Christian Prophec.* p.23

Author 2 - who wrote Similitudes 1 to 7.

Author 3 - wrote Vision 5 as an introduction to the Mandates and joined Mandates and Similitudes by adding Mandate 12.3.3 to 12.6.5.

Author 4 - added Visions 1 to 4.

Author 5 - wrote and added Similitude 8.

Author 6 - concluded the work by writing and adding Similitude 9.

Similitude 10 is not included in the analysis because it exists in Latin rather than Greek.

Unfortunately, Coleborne's theory was rejected as being too complex and his work has never been critically analysed in order to discover its success or failure. His approach, however, was novel and all of the tedious work done without the aid of a computer! If his approach was brilliant, it was matched only by the brilliance of its failure. The reasons for the failure need to be discovered and recorded. This will be done, in due course, as part of the work of this thesis.

Inspired by Coleborne's efforts we now examine whether modern multivariate statistical methods and computer processing power can help in solving the question of the number of authors? This will, in turn, assist in developing a conclusion for the date of writing of The Shepherd. Only then will we be in a position to understand the social location of the Hermas community by examination of the internal evidence of the writing. Before we begin that journey, the next chapter will deal with the background to, and the development of, a modern approach to statistical authorship attribution studies.

62 Coleborne's work has frequently been cited and dismissed without much discussion.
One of the goals of literary criticism has been to find a method by which the authorship of a text could be determined. In an age of deconstructionism and cynicism beliefs are questioned and challenged with alternative theories designed to turn the established ideas on their heads. Everything is challenged. Not just religious belief or the authorship of cherished texts, but the whole 'establishment' is regarded as corrupted. Conspiracy theories attain a new vogue in such a climate, and many people are prepared to believe a 'cover-up' with regard to the church authorities and the fictional “Da Vinci Code”, or the American Government and the reality of the Apollo moon-landings. Properly constructed academic criticism does have an important part to play in the advance of knowledge, and such a reasoned approach has been seen in the realm of literature.

Some Shakespearian works have always been the subject of debate as to their true author, but the debate extends to other authors too, some close to our own living memory. For example, L.F. Baum is well known for his writing of *The Wizard of Oz*; based on his success with this story he produced at least 14 other books in a series. However, some critics argue that the 15th book in the series (*The Royal Book of Oz*) published in 1921, was not written by Baum, but by R.P. Thompson. There is strong circumstantial evidence that this was the case, the driving force for a change of author having been the untimely death of Baum in 1919 and the commercial interest of the publisher in completing the series.

This dispute over *The Royal Book of Oz* has occurred in the very recent past, and there are still extant personal and business documents relating to the events leading up to publication of the work. It is much more difficult to argue against the authorship of a work written several hundred years ago. But, the Oz literature is an ideal candidate for statistical authorship studies since any conclusions obtained by a given methodology can be checked by reference to extant documents. This was indeed the case, when in 2003 the corpus was investigated statistically by
J.N.Binongo. The results confirmed that Thompson had indeed written the 15th book in the series.

If there are no other extant documents relating to the issue, then it is only the work itself that provides the clues in the dispute. In statistical analysis this requires good methodology. Indeed, the situation becomes even more complex when we are concerned with documents written in the 1st and 2nd centuries AD, because this is the era when material was copied by hand. Consequently, the extant documents are copies of copies and may therefore contain additional text added by the copyists, either intentionally or accidentally. Over a period of time pages may have been damaged, lost or even replaced out of sequence, raising questions about the possibility of determining the authorship of an ancient corpus of texts such as the Pauline or Ignatian letters. This chapter deals with how this problem has been tackled in recent times and the methods that may be used to determine single or multiple authorship of The Shepherd of Hermas.

The idea that a written text will contain linguistic clues to the author of that text is not new. Jean Astruc (1753) could be said to have begun the quest when he postulated two sources for the Book of Genesis, based on the two different Hebrew words used for 'God'. He designated these two sources as J and E. Modification and development of this method led to the 'Documentary Hypothesis' of the Pentateuch and, arguably, was the beginnings of Textual Criticism as a scientific discipline. But, Astruc's method is precisely that followed by statistical linguists ever since, namely:

1. Find a measure that characterises the author of the text (in Astruc it was the divine name).
2. Tabulate the frequencies of the occurrences of the measure in the texts to be studied.
3. Group those texts which are most similar (in Astruc two groups, J and E).
4. Explain similarities and differences ie. do they form patterns that can sensibly be attributed to different authors?

Hidden within this methodological framework are two major assumptions that cannot pass discussion:

a) That grammatical construction can be used to indicate an author. In the above example
the difference of a single word was chosen to distinguish authorship. This leads to the
second assumption that:

b) The grammatical writing style of an author remains constant throughout the period of
writing a text. In the above example it means that because there are two different words
used for the term 'God', then there are two different authors and not a single author who
uses the name interchangeably.

These assumptions will be tested and evaluated in the chapters that follow.
First it is necessary to say a few words about style, however, and whether there is
such a thing as authorship style.

**Grammatical Construction and Authorship Style**

Style is a very difficult concept to define\(^{64}\) and usually consists merely of a list
of features that constitute a particular style, such as, epistolary style, formal style,
conversational style etc. Labov\(^{65}\) attempted to formalize style so that it could be
measured and defined. He states,

*The most important problem to be solved in the attack on sociolinguistic structure is the quantification of the dimension of style.*

Unfortunately, he concluded that the numbers of variables were so great that
the task of quantification was an impossible one! D.D. Clarke\(^{66}\) also had little
success when he tried to develop a set of rules of syntax and grammar for the
analysis of conversational style. He too found the task to be an impossible one.

The complexity of the problem is obvious. A text is designed to be an
interaction of the author with an audience. While the author writes, the author has a
mental idea of the audience that he seeks to inform, persuade or encourage. The
intended audience and the need for clarity determine the author's literary manner,
choice of words and phrases. The author needs to adapt his/her own language

---

64 Freeborn, D. *Style Text Analysis and Linguistic Criticism.* Chapters 1-3.
65 Labov, W. *Sociolinguistic Patterns.* p.245
66 Clarke D.D. *Language And action.*
expression to the social context of the intended audience. An author may therefore change the style of writing between different texts, and also within a text.

When examining style, the factors which are usually taken into account, are, lexis and syntax. Lexical patterns include use of abstract terms, formal language, neutral or emotive words. Syntactical patterns include simple or complex sentences, use of statements and questions, and adverbial and relative clauses. This approach to style studies is not without critics. Botha\textsuperscript{67}, for example, argues that by only taking the "\textit{minute aspects of (grammatical) style, in isolation}" the historical context and purpose of the text is ignored. Therefore, there is a danger of losing sight of the communicative intention of a text when it is reduced to merely grammatical components in order to obtain "\textit{the quantification of the dimension of style}". This is an important point that cannot be stressed enough, \textit{quantification methods do not provide the final, or only, solution in an analysis; context and common sense are equally as important.}

\textbf{Lexical Richness as a Measure of Authorial Style}

Is there such a thing as measurable authorial style? If so, what are the valid markers of that style? Are they reliable and reproducible from text to text or author to author? These are basic questions, but they have not been well discussed in the scholarly literature of this subject. Those who have claimed good results have chosen the markers empirically from the data without any grammatical theory that underpins their choice. Biber\textsuperscript{68} attempted to discover if there was any correlation between genre and linguistic features. Taking 67 linguistic features he was able, by factor analysis, to uncover 7 “dimensions of register variation” which seemed to have a correlation with genre. Chaski\textsuperscript{69} examined the use of empirical markers and tested them by a number of blind trials. The results were not impressive, although she used small sample sizes in an attempt to develop a forensically secure methodology that would stand as expert evidence in American legal systems. However, it is generally agreed that unless, or until, there is also a valid grammatical theory that supports and explains the inclusion of markers of authorship, a forensically reliable methodology

\begin{flushleft}
\footnotesize
\begin{itemize}
\item \textsuperscript{68} Biber, D. \textit{Variations Across Speech and Writing.}
\item \textsuperscript{69} Chaski, C.E. “Empirical Evaluations of Language-Based Author Identification Techniques.” \textit{Forensic Linguistics}, (2001) pp.8ff
\end{itemize}
\end{flushleft}
will not exist.

In order to understand the present-day thinking on this subject, it is necessary to give a brief indication of past developments and failures.

We can define authorial style as that part of the writing process where the author does not put much conscious thought into the writing process. It is therefore divorced from the content of the text. The writer will consciously apply other forms of style in the production of text, so that, for example, the writer may apply a formal style in writing a letter to a bank or business, or an emotive style in writing a love poem etc. It is the unconscious authorial style that will lead to authorship identification and not the descriptive content of the text. Authorial style is therefore an indicator of the unconscious psychological thinking of the author. Language is therefore an important diagnostic tool in the evaluation of psychiatric disorder. Grisso70 observed a number of patients and concluded:

Language is an extremely diversified behaviour. Nevertheless, individuals may use language in highly individualistic ways independently of the topics communicated. The results suggest that linguistic styles identify personalities.

We can therefore be confident that authorial style exists. It can therefore be treated, methodologically, in the same way as any other aspect of style. Consequently, lexis and syntax would be used as measures of authorial style. The problem for the researcher becomes that of the choice of measure for this form of style. We noted above that Astruc chose the divine name as a measure of authorial style; a different divine name being equivalent to a different author. This is a valid statistical measure, but it does not have sufficient discriminatory power to be able to distinguish between authors.

One of the early methods for discovering authorship style used sentence length as an authorship marker71. It was thought that some authors would be prone to use long 'wordy' sentences, whereas other authors would be more concise. Therefore, measuring sentence lengths, obtaining the mean and comparing the mean values should indicate differences in authorship style. When this approach was taken with New Testament texts there were special problems to consider. The original Greek

70 Grisso, J.T. “Verbal Behaviour and the Action Through Dimension.” Journal of Abnormal and Social Psychology (1970) 76 pp.265-269. Similarly Hogben, G.H. “Linguistic Style and Personality.” Language and Style (1977) 10 pp270-275 argues that speakers have linguistic habits or traits which are related to their personalities. This study is known as Psycholinguistics.
texts would not have contained punctuation marks; these were added much later and represent an editor's choice. Nevertheless, major stops (full stop, colon and semi-colon) were considered as sentence delimiters. While the advantage of this method was its ease of calculation, the flaws in the method are obvious, as a quick (small) example shows. The number of words/sentence taken from randomly selected sentences from Paul’s letter to the Galatians and from John’s Gospel is recorded in the table below:

<table>
<thead>
<tr>
<th>Sentences</th>
<th>Galatians</th>
<th>John</th>
</tr>
</thead>
<tbody>
<tr>
<td>75</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>54</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>8</td>
<td></td>
</tr>
</tbody>
</table>

Mean 22.8  13.6  
STD 19.2  8.1  
t= 1.73

At first sight it seems that there may be some distinguishing power in using sentence lengths. We can see in the first sentence, randomly chosen, that Paul constructs a sentence 75 words long while none of the sentences in John is greater than 39 words in length. However, if the small random selection of sentences is representative of the whole text, we can use statistics to predict the length of any other sentence randomly selected from either text. Referring back to the table, the mean sentence length and the STD (standard deviation) of the sentences are calculated. Despite a sentence of 75 words, the average sentence length of Paul in Galatians is 23 words (22.75 rounded up to the nearest word) and 14 for John's Gospel. The standard deviation is a measure of the spread of the sentence lengths and is used to predict an expected sentence length of any other sentence taken from the texts. The value predicted would depend on how certain of the result you want to be. The usual level of certainty for many statistical-based inferences is 95% and
corresponds to a value that is 1.96 times the standard deviation either side of the mean value. Therefore, taking any sentence at random from Galatians, we are able to predict that 95 out of 100 of those sentences will contain between -15 and +51 words, while from John it would be -2 to +30 words. Not only do we get a ridiculous result (how can a sentence contain a negative number of words!), but also it clearly shows that many of the sentences from John would fall in the same range as those from Galatians but there would be some sentences that could be ascribed to Galatians because of their longer length. Plotting the predicted density of sentence lengths, we can show, graphically, the nature of the problem.

This diagram shows that sentences in the green shaded area are the only ones that
can confidently be predicted as coming from Galatians, that is, sentences that contain more than 45 words. All other sentences cannot be assigned at all.

We can conclude from all of this that although it is true that some sentences in Galatians are considerably longer than those in John, sentence length alone does not give sufficient discriminatory power to confidently predict differences in authorship style. (I have dealt with this, at this point in the study, in a somewhat simplistic way. It will be necessary to review these concepts in a more detailed fashion in the next chapter.)

Further attempts were made to improve the discriminatory power that such an approach yielded. Consequently sentence length and all words beginning with a vowel letter were chosen as authorship markers. Also a technique taken from quality control of industrial processes was adapted for authorship markers. Again, these attempts suffered the same fate – they lacked discriminatory power.

While Morton and his associates refined the sentence length model, others were tackling the problem from a lexical perspective and using a vocabulary size model. If a text is 5,000 words long it was said to have 5,000 tokens. But a number of these words will be repeated throughout the text and there may only be 2,000 different words in the text. These different words were referred to as types. The type/token ratio is a measure of lexical richness. Mathematically the type token ratio (TTR) is given by:

\[
TTR = \frac{V(N)}{N}
\]

Where \(V(N)\) is the number of types

\(N\) is the number of tokens.

In the example given above, \(TTR = 2000/5000\)

That is a type token ratio of 0.4

---

72 Morton, A. Q. & Michealson, S. The Qsum Plot
Good suggested that the TTR would be constant for a given authorial style (AS) and that the differences between authors would be revealed as different TTR values. Thus:

\[
\text{AS}(i) = f(TTR(i)) = \text{constant}
\]

Where \( i \) represents author.

The table below shows the TTR for a number of New Testament and Old Testament texts.

<table>
<thead>
<tr>
<th>Text</th>
<th>TTR</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Galatians</td>
<td>0.3021</td>
<td>2230</td>
</tr>
<tr>
<td>John</td>
<td>0.1778</td>
<td>15641</td>
</tr>
<tr>
<td>Romans</td>
<td>0.284</td>
<td>7109</td>
</tr>
<tr>
<td>Genesis</td>
<td>0.1475</td>
<td>32553</td>
</tr>
<tr>
<td>Acts</td>
<td>0.2858</td>
<td>18379</td>
</tr>
<tr>
<td>Revelation</td>
<td>0.2175</td>
<td>9841</td>
</tr>
<tr>
<td>Exodus</td>
<td>0.167</td>
<td>24818</td>
</tr>
<tr>
<td>Mark</td>
<td>0.2602</td>
<td>11313</td>
</tr>
<tr>
<td>1 Corinthians</td>
<td>0.291</td>
<td>6829</td>
</tr>
<tr>
<td>Jeremiah</td>
<td>0.1839</td>
<td>28946</td>
</tr>
</tbody>
</table>

Galatians and John seem to have very different TTR values, whereas Galatians, Romans and 1 Corinthians (a Pauline corpus) seem to be closer in value. These results seem to be reasonably impressive, until a graph is plotted showing the variation of TTR with text size \( N \).

---

The graph shows that the TTR is a function of text length and is therefore NOT constant with respect to author. Therefore, TTR alone does not have the discriminatory power to act as an authorship marker.

Over the next twenty years a number of alternative measures (of varying complexity) were proposed. These were all variations on the theme of lexical richness designed to remove the vocabulary size-text size dependence. They include the suggestions summarised below (an extended discussion can be found in the paper by Tweedie and Baayen)\textsuperscript{74}.

Guiraud\textsuperscript{75} modified the TTR using the square root of text size:

\[ R = \frac{V(\sqrt{N})}{\sqrt{N}} \]

Herdan\textsuperscript{76} proposed a simple power function of text size:

---


\textsuperscript{75} Guiraud H. \textit{Les Caracteres Statistiques du Vocabulaire}.

\textsuperscript{76} Herdan G. \textit{Quantitative Linguistics}.
\[ V(N) = N^c \]

where

\[ C = \frac{\log (V(N))}{\log (N)} \]

Dugast\textsuperscript{77} suggested a simple power function of log of text size:

\[ V(N) = \log^k(N) \]

where

\[ k = \frac{\log(V(N))}{\log(\log(N))} \]

Bruner\textsuperscript{78} changed the base of the simple power function of the log of a power of the text size!

\[ V(N) = \log_a^w(W) \]

where

\[ W = N^{\alpha(N)} \]

and \( \alpha \) is taken as 0.172.

Yet no matter how complex the mathematical expressions became, the conclusion was always the same, namely, the measure of lexical richness could not discriminate authorial style, although Pruscha continued to look for a suitable function\textsuperscript{79}.

It was clear that sentence length and lexical richness were not authorial

\textsuperscript{77} Dugast D. \textit{Vocabulaire et Stylistique}.
\textsuperscript{78} BrunetE. \textit{Vocabulaire de Jean Giraudoux}.
markers able to distinguish authorship. A radical change in thinking was required. Burrows\textsuperscript{80}, working on an idea first proposed by Mosteller and Wallace in 1964, pioneered the use of function words that were seen as important sentence components in the study of generative grammar. We now look at this idea in a little more detail.

**Function Words as Authorial Markers.**

Advocates of function word measurements argue that these are words that are used more or less unconsciously by an author and that they are therefore representative of the syntactic and stylistic habit of that author. These are bold claims, but what is the theory behind function words?

Generative grammar became popular in the 1980's, the time when Burrows was working on his proposal for a measure of authorial style. The ideas of a generative grammar were developed by Chomsky and his quest for a 'Universal Grammar'. Generative grammar generates, or specifies, how relevant sentence structures are formed. The principles apply to a large number of languages other than English, so that it is possible for students to generate sentences in other languages, even though they had limited understanding of the syntax of the other language. Because generative grammar required the analysis of the constituent parts of a sentence, ego noun phrase, prepositional phrase etc., words that had been considered of minor importance in standard grammar now took on a new and important role.

Words were divided into two types, namely,

1. Open class words.
2. Closed class words.

Open class words consist of those words belonging to the lexical category that can freely add new words to the set. They are words that are nouns, verbs, adjectives, which tend to supply the bulk of the meaning to a sentence. Consequently they are sometimes called 'contentives'. Closed class words are those that give no descriptive content to a sentence but serve a grammatical function only. They include words belonging to determiner, auxiliary verb, conjunction, and degree adverb categories. Because they have grammatical function they are sometimes called

functors’, or ‘function words’.

In English there are no hard and fast boundaries between these two classes.

Some prepositions have lexical meaning, such as location (behind) and direction (toward); others have little meaning (of or to). Many are used to introduce sentences (after, for, like) and are therefore similar to prototypical function words.  

Radford gives a simple test to distinguish between contentives and functors. He suggested that if a word has an antonym it is a contentive. Particles, auxiliaries, determiners, pronouns and complementizers are functors. Perhaps analysis of a short passage of text will clarify some issues. Consider the following:

When, therefore, I ceased asking her all these things, she said to me: "Would you like to see something else?" I was anxious to see it, and rejoiced greatly at the prospect. She looked at me and smiled and said to me: "Do you see seven women round the tower?" "Yes", I said; "I see them." "This tower is being supported by them according to the commandment of the Lord."

(V3.8.1-2)

We could summarize this passage: "Seven women supporting the tower", or, we may be more specific and state: "Hermas' vision of seven women supporting the tower".

In the first statement, "Seven women supporting the tower", we have simply taken the words from the text which give the main point to the passage. All of the other sentences and phrases in this passage are informational, they could be changed to convey other information, but as long as the words - "Do you see seven women round the tower? This tower is being supported by them." - are not changed, then the theme of the passage remains clear. To test whether this is true consider the following changes to the passage:

While I was eating my lunch she asked me: "would you like some pudding?" I nodded positively. She whispered and said to me: "Do you see seven women round the tower?" I took a sip of coffee. "This tower is being supported by them".

Although the impact is lessened, the theme remains the same, that is, the tower.

Why, then, are we drawn to this conclusion? It cannot be because "Do you
When, therefore, I ceased asking her all these things, she said to me: "Would you like to see something else?" I was anxious to see it, and rejoiced greatly at the prospect. She looked at me and smiled and said to me: "Do you see seven women round the tower?" "Yes", I said; "I see them." "This tower is being supported by them according to the commandment of the Lord."

The idea of seeing is important, the word see (highlighted in yellow) occurring four times in the passage. There is also a related word (not highlighted) to look. Next, the seven women (highlighted in green) occur three times and tower (highlighted in blue) twice.

But as we note from our radically changed story, we can change the initial stages of the story, but the repetition of the word 'tower' with its transformation to subject is still strong enough to carry the meaning of the passage. Even without the initial question, "would you like to see" / "anxious to see" preliminaries although the force of the passage is considerably weakened the theme remains clear.

Thus, the meaning of the paragraph is conveyed by the key words 'see', 'tower' and 'them' = 'seven women', and we can summarize the passage with the statement, "Seven women supporting the tower".

The second summary statement, "Hermas' vision of seven women supporting the tower", adds further information that is the product of the author of the statement. The key word 'see' is introduced as the word 'vision' based on the author's understanding of the events in the narrative that precedes this passage. The information that the subject of the vision is 'Hermas' is added for clarity and a simple "A vision of seven women supporting the tower" would suffice.

Finally, the key words 'seven women' and 'tower' could be replaced by other
nouns, say, 'seven ropes' and 'boat'. The passage still makes sense but the theme changes. Even the quantifier of the first noun can be changed or dropped, so that all we are left with are the two nouns.

We can therefore clearly observe that only a few nouns and verbs are required to carry the meaning of this passage. These categories, nouns and verbs, are the open class words or, contentives and are narrative dependent, not author dependent. However, we can see too that the author has carefully constructed the text in such a way that it interacts with the reader on a sub-conscious level, as well as on a cognitive level.

If we now highlight the function words, we can begin to understand the part they play in this short passage:

\begin{verbatim}
When, therefore, I ceased asking yet all these things, she said to me: "Would you like to see something else?" I was anxious to see it, and rejoiced greatly at the prospect. She looked at me and smiled and said to me: "Do you see seven women round the tower?" "Yes", I said: "I see them." "This tower is being supported by them according to the commandment of the Lord."
\end{verbatim}

The personal pronouns encode the grammatical property of person, number, gender and case, but they have no descriptive content. They have their scope of reference determined by the context of the passage in which they occur. The I that had ceased asking in the above passage is known only from what has preceded, i.e. the I refers to Hermas, but in other parts of the text the person speaking might be different, the Shepherd for example, but the same word, I, might be used. Technically, then, personal pronouns are functors. Longobardi\(^3\) argues that pronouns belong to the same class as other determiners. The author has few choices in their use within the text, other than the frequency with which they are used. However, because they are pronouns that stand in place of nouns and can therefore be identified, I am not going to include them in the class of function words used in this study. They will enter the analysis in relation to thematic consideration.

Determiners also encode the grammatical property of person, number, gender and case, but they have no descriptive content. This class contains words such as the definite article and 'this' and 'that', which tend to introduce referring expressions. The 'the' in the expression 'the tower' refers to a specific tower assumed to be known to the hearer. However, the strong association of the definite article with noun and noun

phrases will probably have a negative effect in determining authorial style. Why? Because its pervasive and wide range of use will swamp the authorial style. Therefore, they are not included in my list of function markers for this study.

Particles and conjunctions are also function words. Particles are those words "which are invariable in form and which don't fit easily into traditional systems of grammatical categories." Included in this category are words such as 'to', the infinitival particle, and 'not' the negative particle. Similarly coordinating conjunctions and subordinating conjunctions are of invariable form. Although these conjunctions do have a close relationship with noun phrases they can give important indications of authorial style. An over frequent use of conjunctions can be found in the writings of young children, and also in some cases of explanatory conversations. These will be included in the list of function words used in this study.

Adverbial forms, such as 'therefore', are logical conjunctions in Greek and are therefore included in the list of function words used in this study. Indeed, all adverbs are included in the list. This may be over-kill, but statistical selection techniques (described later) will decide whether they will be used in subsequent analysis.

Technically, auxiliary verbs are included in the class of function words because they behave differently to other verbs. In particular, auxiliary verbs take a verb expression as their complement. Auxiliaries, unlike other verbs, can be directly negated and usually contracted eg. "wouldn't", "isn't" etc. Another function of auxiliaries can be seen in the passage of text above, that is, they can undergo inversion to form a question ie. "you would.." inverts to "would you..?" Other verbs require the dummy auxiliary "do". Questions are very important in The Shepherd of Hermas, but are they indicative of authorial style or genre style? Auxiliary verbs in Greek are not as straightforward as their English counterparts and are not included in the list of function words used in this study.

Thus our short example passage has enabled us to see how function words can be important indicators of authorial style. Baayen et al\textsuperscript{84} used function words successfully in their study of Milton's works. It has since been common practice in studies of English texts to measure 50 to 100 of the most common function words.

\textsuperscript{84} Radford, A. Syntax. p.267
found in the texts of study before performing a multivariate statistical analysis (more about that later).

Conclusion

The failure of early attempts to identify markers of authorial style based upon crude, but easily measured items, led to the possibility of function words being considered good and reliable markers. Empirical studies of English texts have given good results, in that they appear to have discriminatory power. Transferring these ideas to Greek texts is not as straightforward as it seems, but a general rule emerges from the discussion above. The major forms of Greek function words to be used include those Greek words that are invariant in respect to pronomial suffix or case ending. These include particles, conjunctions and adverbs. The author has no choice in their form, only in their frequency of use. A list of such words in Greek can be easily obtained eg. Porter\textsuperscript{86}, or more generally from a corpus of Greek texts. For the sake of this study I have used the works of The Septuagint, Apocrypha, New Testament, The Shepherd, The Didache, Polycarp to the Philippians, Martyrdom of Polycarp, The Epistle to Diognetus, 1 Clement and Barnabas as my corpus of texts, from which I have extracted all of those words which fit the criteria I have previously described. From this ‘Master-set’ it is possible to choose a sub-set of words to be used in any instance.

The greatest need, however, is to identify a reliable, reproducible and accurate working model for authorship studies, and to give a theoretical framework to that model. This is the task of the next few chapters.

86 Porter S.E. Idioms of New Testament Greek. p. 204ff
Model Building

In this thesis, attention is focussed on quantitative statistical analysis. The model building process is one of analysis, refinement and checking, and rarely is a suitable model arrived at without going through a cycle of building.

The first stage involves construction of a mathematical model that may be used to identify authorship of texts. We must use any available theory of authorship and any pertinent data that may be available in order to tentatively define the model. We have already seen how past theories of authorship based on lexical richness has failed to provide an adequate working model, but the use of function words may present a suitable model. There is not a clearly defined theory of authorship attributes that can be used to base this idea in scientific terms, rather, the theory has arisen from the data simply because it appears to work. The work of an author is a complex mixture of psychological and cultural influences. These factors cannot be measured by a single criterion. Rather, psychological testing relies on a broad spectrum of data and categories of behaviour, intelligence etc. are necessarily 'fuzzy' bounded categories. This means that psychological traits are not as easily defined, or explained in more precise terms, such as can be achieved with scientific/theoretical disciplines like astronomical physics. Our building process, therefore, mainly relies on the data which then defines the model. Since the model contains many unknown parameters, it is necessary to estimate their values by applying statistical methods to the data.

The second stage involves an assessment of the adequacy of the fitted model. A model could be found to be unsuitable for a number of reasons; for example it could contain unsuitable or inappropriate variables, or it could model something other than what is intended. This being the case, the model needs to be refined or re-defined until a suitable "working model" is obtained.

The third stage is the testing stage. Having obtained a working model it must then be tested on a new set of data in order to assess the stability and ruggedness of the model under these different circumstances. It can also be tested with specially
constructed data sets that are specifically designed to test the limitations of the model.

If the model passes all of its testing phases only then can it be proposed as a solution to the defined problem. Finally there is an explanation of the results and a proposal of a theory to indicate how and why the model works.

These three phases, and the model building process, can be shown diagrammatically as follows:

This defines a procedure that can be followed for all model-building problems. In our case, the problem can be defined as:

“Finding a set of words that are able to distinguish the psychological process of the writer of a given text, and to show that different writers have different and individual psychological processes.”

Psychological process is different to a writer’s style. A writer may change the style used in a text in order to convey meaning. For example, a verbose, descriptive style may suddenly change to a terse style in order to convey urgency. While this may be a deliberate ploy on the part of the writer, it displays the maturity and skill of the writer’s craft, not the psychology of the person of the writer. Authorship style is a conscious effort of the author to display skill and effect. Psychological process is the subconscious effort of the author using words that do not add descriptive content
or meaning or that display skill or craft. It comprises of words that are written without thought\textsuperscript{87} because they are commonplace. A writer's style may change with the type and purpose of the writing, and, possibly, with the age and experience of the writer, but psychological process remains consistent regardless of style. It remains consistent (not necessarily constant) because there is no conscious effort to improve or modify it.

Theory does not exist to tell us what words should be studied in order to discover a psychological process, except in broad terms to suggest function words. Consequently, the first stages of our model building involve obtaining the data i.e. some measure of the function words used in texts. Before we consider - “What words?” “What texts?” - there is one important final remark to be made concerning method, that is, to explain 'The Principle of Parsimony'. This principle states:

\begin{quote}
In a choice among competing hypotheses, other things being equal, the simplest is preferable.
\end{quote}

The simplest model is chosen because it is easier to explain how the model operates, and to detect any failures or shortcomings in its results. More complex ideas and models may form as a result, but in the first instance simplest is best.

\section*{Data Collection And Manipulation Before Analysis}

\subsection*{a) Data Collection}

In the initial stages of the model building process we have to carefully choose the texts to be studied. These should be a good cross-section of texts representing different authors and styles of writing. Preferably it should include long and short texts, but each text should be selected for a particular purpose in testing the methodology.

The following texts are initially used in the model building stage, and a brief indication of why they were chosen is also given:

\begin{itemize}
  \item \textbf{GENESIS.} This text was chosen because we can claim with absolute
\end{itemize}

\textsuperscript{87} Or are less consciously chosen by the author. This may have an impact on cases of deliberately forged texts. The function words are less likely to be treated with the same importance by the forger as the contentives.
certainty that the author of this text is not the same author as The Shepherd. Any model that concludes otherwise must be measuring something other than authorship. We are also certain that the LXX form of this text is partly, if not wholly, based on the translation of a Hebrew original. Therefore it is also a good source of translation Greek.

- **EXODUS.** All that has been claimed for Genesis applies to Exodus too, but with the addition that we can examine the inter-relatedness of these two texts. Not only will there be possible different authors (the old JEPD), but also there will be the issues relating to the work of editor(s).

- **ROMANS, 1 CORINTHIANS, 2 CORINTHIANS, GALATIANS.** These four texts represent a Pauline Corpus written over an extended period of time. This will test the idea that an author’s psychological process can still be identified from works written at different periods of time. They are also texts written in Greek; consequently Old Testament quotations are removed. These quotations are not the work of the author of these letters. In these texts we have an example of Koine Greek.

- **LUKE, ACTS.** Luke and Acts are considered to be written by the same author, but it is also thought Luke used various ‘sources’ in the construction of the Gospel account. While it is not the purpose of this thesis to examine the possible source material, it does help us to determine the effectiveness of the model to identify editorial material.

- **JOHN’S GOSPEL.** This text will allow a comparison of a text of similar genre to Luke. However, it is clear that John is uniquely different to the other gospel texts. It will allow us to check that the model is able to distinguish author rather than literary style.

- **REVELATION.** If we accept a date for this text towards the close of the 1st century, we have a text that is located within the earliest possible time period in which The Shepherd could have been written. We can check any similarity in expression that may be a function of time. Also, like The Shepherd, Revelation also contains visions of beasts and of angels and may be considered to be of similar literary style.

- **THE BOOK OF WISDOM.** This text was chosen because the first part of it deals with teaching on immorality and Wisdom. These subjects are also
dealt with in The Shepherd, so again we have similarity of material. It is believed to have been written in Greek, and to consist of two parts. The second part (11:5 – 19:22) seems to be of a different literary style to the first part. It will be interesting to discover whether such a dramatic change in literary style is the result of two different authors.

- **ECCLESIASTICUS.** A text on morals that may be of the same genre as parts of The Shepherd. Unlike Wisdom, Ecclesiasticus was written in Hebrew and is therefore a translation text in the Apocrypha. It would be of interest to discover whether the author of The Shepherd was influenced by such a work as this.

- **2 ESDRAS.** This text contains seven visions and may have been written between 81-96AD. It is therefore of a similar time to The Shepherd and was probably written in Greek.

- **THE SHEPHERD OF HERMAS.** Finally, the text of interest. For purposes of analysis, Visions, Mandates and Similitudes are treated as separate independent works. The purpose of the exercise is to discover whether a single author wrote all three sections, or whether different authors wrote or edited different sections.

Each text was divided into separate 3000 consecutive-word blocks. The Shepherd of Hermas is treated as three independent parts, Visions, Mandates and Similitudes since we want to discover the relationship between these parts with regard to authorship and editing.

Where necessary, quotations within texts were removed prior to forming the 3000-word block. This is not as straightforward as it sounds. Borrowed material is embedded into the flow of a writer’s text without necessarily crediting the material88 This was a common phenomenon in antiquity, and in the writings of the Early Church Fathers is a particular problem. Quotations from sources such as the LXX tend to be more paraphrases of the author’s construction and not a direct quotation. This free association means that any attempt by a historian to unravel the quotation tends to be more guess-work than anything else. Therefore, only those parts of texts that are clearly marked by the author as quotation eg. the use of formulaic expression

such as "it is written", "the scripture says" etc. are deemed quotation and removed.

This results in 62 blocks of text, each block consisting of 3000 consecutive words, covering the range of our chosen texts. These blocks of text will hereafter be referred to as cases.

A database was constructed which contains all words of grammatical class preposition, conjunction, particle, interjection and adverb occurring in the LXX, Apocrypha and New Testament (full list in appendix 3.1).

The next step was to count the variables in all of the cases. This can be achieved easily by using any relational database software with SQL (Structured Query Language) capabilities and a little computer programming knowledge. I used Corel Paradox 9, which I was able to program to do the tasks I required, a user input form allowing easy selection of manipulation of the process variables.

The result of following this procedure is a data table consisting of 62 rows of cases and 230 columns of word counts, hereafter referred to as variables.

c) Data Manipulation

The data that has been collected presents a new series of challenges. Some columns of data contain relatively large counts because they are common words, the conjunction καὶ ('and') has a mean count of 230 per 3000 word block. On the other hand, words such as κατεναντίον ('opposite') which functions as an adverb and an improper preposition has a mean count of 0.35 per 3000 word block. That means, of course, the column of κατεναντίον counts will consist of many zero counts, in fact 49 out of the 62 cases will be zero counts. While it may be possible in some instances to infer conclusions from absences, it becomes absurd if there are many instances of zero counts where conclusions will then be formed on the basis of nothing!

Not only is there the logical problem of zero counts, there is also a mathematical problem. There will be a high correlation between columns which contain a high proportion of zero counts, and this leads to the problem of multicollinearity. This occurs when an included variable is so highly correlated with others in the data-set that it is a function of the others. For example, if a study of Egyptian mummies included measuring the skull diameter and it was measured in
inches and in centimetres, clearly these variables are related. Inclusion of the two measurements adds no more information than having just one of them. The resulting matrix would show collinearity and there is a redundancy of variables. In that simple example it would be easy to spot the problem. In larger data-sets the error might not be so easily noticed. With our data, in particular, if there are a large number of zero values associated with variables they might appear to be related in a collinear fashion.

If there is multicollinearity present in a matrix of data, not only will it introduce redundant variables into the model, but, it will affect the determinant of the matrix so that it becomes zero or zero beyond the accuracy of the microprocessor ie. \(9 \times 10^{-36}\). The matrix is thus ill-conditioned and will tend to give peculiar results, particularly on the extremes of the data. Moreover, if any statistical technique requires matrix inversion, then such large numbers are produced that the microprocessor becomes overloaded and causes the computer to crash. It is comforting to know that the problem of logic will ‘blow the memory’ of the computer too!

Generally, the results of any analysis are better (more reliable) if multicollinearity is dealt with at the start before any analysis begins. In order to test for multicollinearity a linear regression algorithm must be run with each variable in turn acting as predictor variable (y) against all the other variables (x). This allows calculation of a multiple R squared coefficient for each variable regressed on the remaining variables. If the mR\(s\) coefficient is greater than 0.9, the two variables are considered to be collinear and one of them can be dropped from the matrix. By backward deletion, forward selection, or a mixture of both, the offending variables can be removed from subsequent analysis.

However, there are yet more problems with the data which are even more serious in nature. Consider the count of the preposition \(\delta\alpha\) (‘through’), whose frequency histogram is shown below. The count can take any value be from zero upward, and in this case the distribution is skewed. With lower count variables the situation is likely to be worse. Before we can begin to tackle the collinearity problem, it would be necessary to find a transformation that would normalize the data.
The histogram shows that the raw data frequency of this word is highly skewed and non-continuous. The solid line is the Sheather-Jones direct plug-in density estimation of this data. The calculation of simple mean and standard deviation would be meaningless.

There are a number of transformations that could be used:

- **Standardization.** The most common method is to standardize the data to zero mean and unit variance. This is the so-called z-standardization given by,

\[ Z_{ij} = \frac{x_{ij} - \mu_j}{\sigma_{ij}} \]

where \( \mu_j \) = mean of column j,

\( \sigma_{ij} \) = standard deviation of values in column j.

(Fig. 3.1)
This requires the mean and standard deviation of all the counts \((x_i)\) in the column \(j\) to be calculated. Remember that a column will contain counts from all the texts of interest, which means that using this transformation for a particular word count will centre the transformed data about its average from all texts. Effectively this compresses the data and causes loss of some of the word count differences between the texts. The use of standard deviation in this context assumes that the data is normal, which may not be a reasonable approximation. This transformation has the effect of giving each column equal importance, so that high-count words are no more important than low-count words. Since the data has been shown to be non-normal, this would not be a good transformation to use on this data.

- Square Root Transformation. It is well known that in the type of data we are using (count data) the variance is proportional to the mean. It is common practice to overcome this variance problem by taking the square root of the data, or, sometimes if the counts are zero or small, \(\sqrt{\text{data}} + \sqrt{\text{(data+1)}}\). Taking the square root has a greater effect on large counts than on small counts and reduces the relative difference between them.

89 Bolton, S. *Pharmaceutical Statistics*. p.353

90 Weisberg, S. *Applied Linear Regression*. 56
The disadvantage is that this involves a certain amount of variable selection because reducing the relative count difference increases the overall importance of the smaller count variables. The question is whether we want to proceed with analysis based on that assumption.

- Logarithmic Transformation. Taking logs of the original data (either base 10 or base e) is commonly used when data is skewed to the right, as in the case of the count data.
Apart from difficulties of interpretation of log data, the log transformation normalizes the variance so that it becomes (approximately) equal to the coefficient of variation (that is the standard deviation divided by the mean). In effect this transforms the data to data that is homogeneous in standard deviation. But we already suspect that the data is not homogeneous, therefore this transformation would not be advisable.

- Arcsine Transformation. This transformation is used when analyzing proportions. Dividing the word counts by the number of words in the text block would give the necessary proportions. This transformation is not advised when the proportions are very small (0.03), or very large (0.97). Most of the count data collected from the texts would give very small proportions and therefore this transformation is not recommended.
While there are several other transformations that are known, the above
discussion is sufficient to indicate the problem. Many of the articles relating to
authorship studies suggest a Z-transformation of the data followed by a Principal
Component analysis, but almost never state that the data is normal. We will return to
this in a later chapter which specifically deals with PCA.

The 62x230 data matrix poses yet one more difficulty; there are more
variables than cases. This will cause difficulties when using most statistical
techniques. However, there is one technique that can be used with skewed variables,
which can handle more variables than cases and is invariant to transformation of
data; that method is “Random forest”. Not only is it an ideal tool for the count data,
but it can also produce a dissimilarity matrix that can be used in the statistical
methods that we are going to use in the following chapters. Because the method is so
useful as a starting point to further analysis, it will be described in the next chapter.
As mentioned above, the commonly used method is PCA, but I have shown that
there needs to be a preliminary treatment of the data first.

Before I begin to describe in more detail the method that fits all of the
difficult criteria that the count data presents, we need to understand some basic
terminology.

**Classification and Discrimination**

Given discrete blocks of texts, whose function word counts we have obtained,
we want to group together those blocks that are similar to one another and
distinguish them from other groups that are not similar. This is a process of
classification/discrimination by which we hope to group together the text blocks that
belong together by virtue of the fact that they were produced by the same author.
Different groups then represent different authors. Therefore, we will need to apply
statistical procedures related to classification/discrimination.

Because of the growing importance of these procedures in all areas of data
analysis, the definitions of classification and discrimination have been refined over
recent years.

**Classification** is now defined as the allocation of cases to groups. Having
specified the groups, new cases are then classified according to the pre-specified
groups. Medical diagnosis is a classification problem in this sense of the definition.
A patient showing symptoms “a”, “b”, “c” is diagnosed as having condition “g”. Similarly, applicants for financial credit can be categorised according to risk – the pre-specified groups being “low”, “medium”, and “high”. A new applicant having a certain annual income, social and marital status etc. can then be assessed as to the level of financial risk and credit given/not given accordingly. Classification thus enables a decision to be made without detailed explanation of the variables measured. There is no need to understand the relationship the variables have with one another, or with the final classification group. Classification simply offers a decision rule by which new cases can then be classified.

**Discrimination**, however, is concerned with the interpretation of the classification. The researcher seeks to understand and explain why one group is classified using particular variables and another group with different variables. Using the medical diagnosis example, the researcher may be more interested as to why a certain disease seems to affect men at an earlier age than women etc. Is this the result of gender difference, or is the disease in fact two distinct diseases? Discrimination, therefore, tries to offers *explanation* in terms of the measured variables.

Our first task, then, is one of classification. We have our data as blocks of text, but we cannot pre-specify the group to which any block belongs. The label information regarding the text from which the block was taken is dropped from the analysis; this is referred to as ‘unsupervised’ (a term borrowed from data mining methods). Our initial task is an exploratory one; can we find any groups in our data? Statistical tools are available to help us uncover the similarities and dissimilarities in our data. All of these tools will produce a graphical display of the results.

**Conclusion**

This chapter has described a basic methodology that can be followed regardless of the texts chosen or the function words used. The type of data and the choice of texts need to be fully and clearly defined at the outset of the analysis. The collected data needs to be examined before using it in any statistical analysis in order to assess the degree of abnormality and any problems that might result from unsuitable use.
The data matrix of function word counts will contain some columns of data having many zero counts on the individual cases. Since this is not an ideal situation because of multicollinearity and ill-conditioned matrices, some researchers might argue the case for simply removing them. By doing this we are imposing our assumption on the data that the larger count function words are more important than the low count ones that we discard. At this stage in the analysis this assumption should not be made.

The assumption that count value is proportional to a variable’s importance in discriminating between possible authorship questions regarding the text block is not resisted solely on logical grounds. Ideally the blocks of text sampled should be made as small as is practically possible in order to assess editorial additions to the text. Typically, editorial amendments will consist of a few words linking together sections of the original text. If the statistical model is able to distinguish authorship of large blocks of text, then the next step would be that of reducing sample size. There will be a cut-off point where the model will break down because the within sample variance will become so great as to compete with between sample variance. But, if the sample size can be reduced to smaller blocks, there will be more of an opportunity of finding editorial additions.

Therefore, it is important not to constrain the analysis by imposing assumptions on the data that may well hinder that analysis. It is important to remember the diagram of the model building process given at the start of this chapter. Part of the model building process is to test the limits of any model that may be proposed, and much of this testing is iterative and involves some repetition. I have tried to keep such repetition to a minimum.

A separate chapter will be devoted to each statistical technique and its assessment in the hope of achieving clarity. The following chapters are also structured in the same format, that is, first, a general and simplified account of the purpose and scope of the statistical method is given. This will enable anyone not interested in the mathematics of the procedure to have enough understanding to be able to engage with the results that are obtained. Second there follows a section which gives some mathematical background in order to indicate any difficulties or improvements that can be made to the methodology. Lastly the results obtained from the analysis are given, together with an assessment of those results.

As previously explained, the next chapter will introduce the method of
Random Forest.
Chapter 4
Unsupervised Random Forest

Introduction

Tree-based models have been in use since 1963 and have been used to classify data. They have been used in a number of disciplines, from the classification of animals in evolutionary theory to social sciences and economic assessments. Their usefulness extends beyond the diagrammatic representation, and tree-based models have been in use under the title of 'Decision Trees'. The fundamental structure of tree-based methods lies in the partitioning of the data according to the application of a series of constraints, until there is no more data to partition.

A simple example will show how the method works. Consider the classification of insects. We might begin the classification by partitioning all of the insects into two groups based upon the constraint of whether or not the insect has wings. Diagrammatically this can be shown as

The first box diagrammatically contains all of the data. The oval represents a node, that is, a place in the tree where a partition takes place. In this instance, the node represents the constraint of insect appearance in terms of wings. There are only two possible outcomes; hence these trees are sometimes called 'binary partition' trees. The initial root (All insect groups) has now been partitioned into two branches. Each branch can be split further into two, based upon another constraint, so that the tree begins to 'grow'.

63
If at each node the answer to the question is yes, then follow the arrow to the right. Only part of the tree is shown here, but it can be seen that eventually the insects cannot be further partitioned. Consequently, the entire genus PHASMIDA is grouped together and branching stops. This is called a leaf node, or terminal node. By ‘growing’ the tree a little more, the entire genus ORTHOPTERA is grouped together. As you can see, a tree may grow many branches before a leaf node is reached, but the leaf node will be the required classification.
The value of such a tree becomes obvious, because it means that someone with little knowledge of insects, if given an unknown insect, will be able to classify it by comparing the unknown with the node questions of the partition tree. Because of this, such trees are sometimes called 'decision trees' because they require decisions to be made at each node, or, 'expert systems' because they act as the expert having the knowledge to guide to a classification.

Expert systems are used in many disciplines, from medical diagnosis, where the nodes might be questions about symptoms, to financial institutions, where the nodes might be questions about tax and economy.

The question now becomes, how do we find the best (optimal) split of the data at each node?

**How Binary Partition Trees Are Constructed**

Because a tree can grow very large, it would be almost impossible to optimize the performance of the whole tree, consequently the split should be optimized at each node only. If we consider our data, which consists of word counts taken from blocks of text taken from Genesis, Exodus, Acts, Luke, 1 Corinthians, 2 Corinthians, Romans, Visions, Mandates, Similitudes, Book of Wisdom, Ecclesiasticus and 2 Esdras i.e. 13 texts, if we now have a sample from one of the above group, we could classify it into any one of the 13 texts with a 92% probability of assigning it to the wrong text by chance alone. In order to improve the chances of a correct classification, we must consider what happens at each node. We need to choose the data split at each node that minimizes the probability of incorrect classification. Breiman suggested a measure that would define the impurity of each node. This measure, based on “Bayesian probabilities” is called the Gini Index.

\[
\text{Gini Index} = 1 - \sum_{i=1}^{I} \pi_i^2
\]

Where \( \pi \) is the prior probability of group \( i \).

Since we do not know the probabilities, they have to be estimated from the proportions of counts in the groups at each node. The binary split which minimizes

---

the Gini index is the one chosen at each stage. Thus the Gini Index for the unsplit (root node) is:

\[ I(\text{root}) = 1 - (n_G^2 + n_E^2 + \ldots) \]

Which, if each has equal probability (1/13),

\[ I(\text{root}) = 0.923 \]

If there is a constraint or value, t, that splits the node, then a number of cases will split down the right branch (say \( n_r \)) and some down the left branch (say \( n_l \)). Knowing the proportions of groups in each branch, the Gini Index can be calculated. The constraint, or value, t which minimizes the index is chosen. Thus the optimal t is found for each node.

Using this procedure\(^\text{92}\) on the raw data in supervised mode, the resulting classification tree can be grown:

\[ \text{92 The algorithm used to do this can be found in rpart: Recursive Partitioning. R package version 3.1-29. S-PLUS 6.x original at: http://mayoresearch.mayo.edu/mayo/research/biostat/splusfunctions.cfm.} \]
An example of a single binary partition tree based on the original data contained in a 62x230 data matrix. The tree is un-pruned.

The root node is first split using the constraint Count $99 \geq 0.5$, that is, if the count of the word \textit{\vland} (a particle that expresses time, 'when', 'the time when') is greater than, or equal to, 0.5 then take the right branch etc. \textit{\vland} is a low count word whose mean is only 0.77, therefore there must be a large number of...
zero counts in its column (44/62 counts are zero). This is a word that would have been removed if we were considering only large count words!

It can be noted that certain texts occur more than once as leaf nodes. Thus 2 Esdras appears twice, indicating that at least two sections of 2 Esdras are somewhat different to one another. Sometimes a leaf node may contain another text that cannot be distinguished. In that case the leaf node is labeled with the text that occurs most in that node.

Note too, only certain variables are necessary in order to produce the tree, only 19 of the 230 original variables are used. Some variables are not important or are not able to be used. These represent the 'noise' in the original data. We therefore have a method which is able to filter out noise. Also, because the method uses proportions, the exact nature of the variable does not matter. This method is invariant to any data transformation, and avoids the problem of finding the 'best' transformation.

What are the disadvantages then? Because the method uses the one-step look ahead approach, there could be many such trees depending on the variable that is chosen at each node! If the procedure was repeated a number of times we could measure the frequency with which each variable was used in tree construction. Breiman proposed a bootstrap method that would do this. By taking a small, randomly selected sub-set of variables from the original data a tree is grown and not pruned. The choice of variables uses sampling with replacement, so that in any given tree construction a variable could be used once, more than once, or not at all. Another sub-set of variables is chosen and a second tree grown. Each time the variables used are stored and counted. If this is performed a large number of times i.e. more than 500 trees grown, the final results give accurate predictions about variables used and groups produced. He termed his method "Random Forest".

Although the above has described a supervised method of growing trees, Random Forest can also be used in an unsupervised mode, that is, the data labels are removed. A proximity matrix and a plot of variable importance can be obtained.

The proximity matrix can be used as data input in clustering programs and multidimensional scaling (next chapter).

A plot of the variable importance is shown below:
The figure shows the relative importance of each variable count used in the analysis. The greater the mean decrease in GINI, the more significant the variable count. The red line shows a suggested cut-off point. Variables below the red line are not of significant importance in discrimination. There are a number of suggestions concerning how the cut-off should be determined, here mean - 3 Standard Errors is used.

Those variables that fall below the red line were not significant in the detection of structure within the data. At this stage, this is not an important consideration since it does not affect the random forest performance. If we want to reduce the data to its significant variables in order to perform Principal Components Analysis, where the
variables cannot exceed the cases, then this will prove to be an important consideration.

**Supervised Random Forest**

Before leaving this chapter a supervised random forest will be performed. In the supervised method the procedure is a leave-one-out method. One of the cases is left out of the tree construction. When the tree has been grown, the case that had been left out of the process is classified according to the newly grown tree. Its classification is recorded. Repeating this process a large number of times will allow a frequency calculation of a case with a classification. A classification error is also computed, which in the forest method is referred to as the ‘Out-of-bag error’ (OOB). If the left-out case is classified into its original class, then the classification error is 0, otherwise it is 1. When the classification error is averaged over a large number of trees, the mean OOB is as good as that obtained if a separate ‘training set’ had been used. Breiman also reports that the OOB error of a Random Forest is usually smaller than the error that is found in conventional methods such as Linear Discriminant Analysis.

Using supervised Random Forest provides a first glimpse of the effectiveness of the proposed model ie. authorship can be detected from function word counts. The classifications are given in the form of a table.

Call:

```r
randomForest(formula = data.dfl[, 1] ~ ., data = data.dfl[, -1], ntree = 2000,
importance = TRUE, proximity = TRUE)
```

Type of random forest: classification

Number of trees: 2000

No. of variables tried at each split: 15

OOB estimate of error rate: 24.62%
Confusion matrix:

```
<table>
<thead>
<tr>
<th></th>
<th>2Es</th>
<th>A</th>
<th>Ex</th>
<th>G</th>
<th>IC</th>
<th>IIC</th>
<th>Jn</th>
<th>Lk</th>
<th>M</th>
<th>Ro</th>
<th>Rv</th>
<th>S</th>
<th>Si</th>
<th>So</th>
<th>V</th>
<th>class.error</th>
</tr>
</thead>
<tbody>
<tr>
<td>2Es</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.000</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.200</td>
</tr>
<tr>
<td>Ex</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.500</td>
</tr>
<tr>
<td>G</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.100</td>
</tr>
<tr>
<td>IC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.500</td>
</tr>
<tr>
<td>IIC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.000</td>
</tr>
<tr>
<td>Jn</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.167</td>
</tr>
<tr>
<td>Lk</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.167</td>
</tr>
<tr>
<td>M</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.000</td>
</tr>
<tr>
<td>Ro</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.500</td>
</tr>
<tr>
<td>Rv</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.000</td>
</tr>
<tr>
<td>S</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.000</td>
</tr>
<tr>
<td>Si</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0.500</td>
</tr>
<tr>
<td>So</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1.000</td>
</tr>
</tbody>
</table>
```

Reading the table horizontally shows that all 4 cases (text blocks) of 2 Esdras are classified as 2 Esdras (i.e. in the 2 Esdras column); hence the class error is 0. Acts falls into two sections, 4 cases are classified as Acts and one case as Luke, thus 1 case out of 5 is considered to be incorrectly classified, hence class error of 0.2. Exodus falls into two parts, 5 cases are classified as Exodus, but 3 cases are classified as Genesis. With Genesis, 1 case is classified as Exodus and 9 cases as Genesis. This suggests one of two things. Either the model and method is not able to classify correctly, in which case the class error is correct, or, the model and method is working well and has discovered at least two authors involved in the Genesis/Exodus text.

Next consider the Pauline Corpus of texts. 1 Corinthians is split into two groups. One case is classified as 1 Corinthians and one case as Romans. This suggests that at least part of 1 Corinthians is too alike to Romans to distinguish between them. 2 Corinthians gets classified with Romans. One part of Romans is classified as 1 Corinthians and one part classified as Romans. This might suggest that parts of 1 Corinthians and Romans are by a different author to 2 Corinthians and the remaining parts of Romans and 1 Corinthians. This suggests two different hands in the production of a Pauline corpus!

The Shepherd Corpus has Mandates, Visions and most of Similitudes in the Similitude column. One case of Similitudes is classified as Mandates. The majority of the corpus is thus classified together with only one other case in a related column.

Sirach is classified into one column, but The Book of Wisdom (So) has part classified with Sirach and part with Wisdom. One of the objectives in including Wisdom was to see if the model could detect the two parts of the text.

Random Forest Terms and Definitions

A tree, \( t \), is constructed by generating a random vector (sub-set of variables) \( t_k \). \( t_k \) is independent of any previous \( t \)'s but has the same distribution. If \( t_k \) is grown using a training set where the input vector is \( x \), the resultant classifier is \( \eta(x, t_k) \).

A random forest is a classifier consisting of a collection of tree structured classifiers \( \{ \eta(x, t_k), k = 1 \ldots N \} \) where \( N \) is the number of examples in the training set, and where the \( t_k \) are independent identically distributed random vectors.

Each tree casts a unit vote for the most popular class at input \( x \).

The accuracy of the random forest depends on the strength of the individual classifiers and the dependence between them. When there are a large number of variables, each individual classifier has little accuracy, but when large numbers of trees are grown the accuracy is improved. The 'margin' measures the extent to which the average number of votes for the right class (remember, it is a training sample) exceeds the average vote for any other class. Therefore, the larger the margin, the more confidence in the classification. The classification error tends to an upper bound limit as the number of trees increases.

The error depends on two factors, a) the strength of the classifiers and b) the correlation between the classifiers. In fact, Breiman shows that the Prediction Error (PE) is related to the strength of the classifier \( s \) by:

\[
PE \leq \frac{\text{var}(M)}{s^2}
\]
that is, the upper bound error is the ratio of the variance of the margin divided by the square of the strength of the classifier.
Where a training set is not used, the ‘leave-one-out’ method produces errors that are equally as good as if a training set were used

Conclusion

The Random Forest Method provides a good start to the analysis because of its ability to deal with the problems of the data. The unsupervised method produces a proximity matrix that can be used as input data in many other conventional statistical methods. These will act as useful validation tools because the RF proximity can then be compared with the standard Euclidean distance proximity of the transformed data.

The supervised RF has provided a first glimpse of the model and data. While it suggests possibilities, it also presents problems. Because the supervised method tries to find variables that will allow the data to be partitioned into the preset groups (texts) it has the tendency to over-cluster the data. It will attempt for instance, to find the variables that will assign Romans, 1 Corinthians, 2 Corinthians and Galatians into 4 separate groups. However, because it is the same author that has produced the texts, they should be clustered to one group. Because the algorithm tries to produce the separate clusters, it sometimes can only group texts by closest similarity. Instead of one group we may therefore have two, hence over-clustering. This does not invalidate the results, rather it helps to uncover the ‘finer’ differences in the texts.

We can see that this has occurred with The Shepherd data. Most of the cases fall into one cluster, but one section of the Visions is assigned to a cluster of its own. Similarly, one section of The Similitudes is assigned to its own cluster. If all other analyses assign The Shepherd into a single cluster, then we can see that the supervised forest algorithm is over-clustering. Although there might be a single author, one part of the Visions is slightly different to the other part of the Visions.

The next chapter will look at clustering methods using both the RF proximity and the standard Euclidean distance measure.
Chapter 5
Cluster Analysis

Introduction

The aim of cluster analysis is to discover groupings amongst the cases in the data matrix. The process is conceptually easy to understand, but the practical considerations are not entirely straightforward. Clusters are found on the basis that those cases which are most similar, in terms of the variables, form clusters. Each time the nearest case is added to the cluster, the within-cluster sum of squares will increase. If the cases to be added are very similar the within-cluster sum of squares will not change significantly. However, as the cases become more separated from one another so the within-cluster variance will increase. Therefore, the objective at each stage is to minimize the increase in within cluster sum of squares.

There are two ways to perform the clustering process:

- Hierarchical clustering.
- Nonhierarchical clustering.

Hierarchical Clustering

In hierarchical methods the data is clustered by a nested series partitioning of the data. This can be performed in two different ways.

Divisive methods begin with all the cases classified into one large cluster. The single cluster is then split into two clusters, the clusters being chosen so that the within-cluster sum of squares is minimum for each cluster. These clusters are then successively split until all of the data is reduced to the individual cases. This is similar to the binary tree partitions of the previous chapter, but in that case the clusters were split on the basis of a single variable value, in this case they are split on a sum of squares criterion.

Agglomerative methods begin with the individuals, clustering together those individuals that are most similar. Small clusters are then agglomerated to produce larger clusters. The clustering criterion is again based on the within-cluster sum of squares (there are other methods of estimating similarity, but for the moment
attention is confined to the sum squares criterion). Clusters are gradually agglomerated until there is one cluster containing all the data.

Unfortunately, the two methods will often not produce the same results. Both suffer from the disadvantage that once a fusion or a split has been made it is irrevocable and the process simply continues to its end.

Hierarchical classification can be represented diagrammatically by a dendrogram.

**Cluster Dendrogram**

![Cluster Dendrogram](image)

*Fig. 5.1 A stylised dendrogram showing the structure of nodes and branches.*

The dendrogram above shows the familiar branch and leaf structure, however, each node represents an agglomeration or split, depending on whether an agglomerative or divisive method is used. The length of the branch is proportional to the agglomeration coefficient, whether it is height or variance.

Assuming an agglomerative technique was used, we can see that cases G and H were the most similar (smallest branches) and so were combined together. A and B were the next cases combined. Case F was close to the newly combined G,H grouping, so it forms a cluster of F,G,H. Eventually two clusters are formed, cluster one containing A,B,C,D and cluster two containing E,F,G,H. These then combine
into a single cluster containing all the data. The length of the branches shows that the two clusters were distinct and well-formed.

The explanation would be reversed if the dendrogram was formed by a divisive technique, so that the single cluster was first split into two well separated clusters etc.

**Stopping Rules**

One of the difficulties of hierarchical clustering is knowing when to stop the process of agglomeration or divisive partition. In order to make the decision about how many clusters are found, there has to be an empirical rather than subjective estimation. For example, in the stylized dendrogram above are there two clusters, three clusters, or four clusters?

Most stopping rules seem to be based on the rate of change of clustering criterion. When individuals are clustered (or formed by splitting), the variance is not large, therefore the rate of change is small. When two separate clusters are combined, the rate of change of variance will rise, depending on how well separated the clusters were. Large increases indicate greater separation of combining clusters. A typical plot of rate of change of agglomeration may look something like:
As the clusters agglomerate from 30 clusters to 10, the rate of change of agglomeration coefficient shows a gentle rise. The trend-line is shown in red, and although there are occasional peaks and troughs, the changes are not statistically significant (the 3*sigma limits could be plotted if necessary). The sudden rise in relative coefficient at cluster 5 is statistically significant, suggesting that clusters 5 and 4 were well separated. There then follows another gradual rise until the significant increase in going from 2 clusters to 1 cluster. We would therefore conclude that the data consists of 2 or 5 clusters.

Another stopping rule is based on Beale’s pseudo-F statistic\(^93\) given by:

\[
P' = \frac{(W_2 - W_1)^* (N - C_1)^* k_1}{W_1^* [(N - C_2)^* k_2 - (N - C_1)^* k_1]}
\]

where there are two possible clustering solutions \(C_1\) and \(C_2\) with cluster 1 being the larger of the two; \(N\) is the number of cases; \(k_1 = C_1^{-2/p}\) and \(k_2 = C_2^{-2/p}\); \(p\) is the number of variables;

\(^93\) See Johnson, D.E. *Applied Multivariate Methods*. p.328
\[ W_1 = \sum_{r=1}^{q_r} \sum_{q=1}^{q_r} (x_{rq} - \bar{x}_r)^T (x_{rq} - \bar{x}_r) \]

If \( F^* \) is greater than the F critical point with \((N - C_2)k_2 - (N - C_1)k_1\) degrees of freedom in the numerator and \((N - C_1)k_1\) in the denominator, then the solution with more clusters is favoured over the solution with fewer clusters.

Using a mixture of the simple plot and the calculation of Beale’s pseudo-F it is possible to arrive at a decision about the number of clusters.

**Dissimilarity Matrix**

We have now to return to the problem of the data. The previous chapter has indicated how we can overcome the difficulty by using the data matrix from the unsupervised Random Forest. The usual method of creating the dissimilarity matrix for clustering is by taking the Euclidean distance between pairs of points. At this stage we will compare results obtained from the Random Forest proximity matrix, the Euclidean matrix and the data transformed by the \( \sqrt{\text{count}} + \sqrt{\text{count} + 1} \) transformation.

**Results**

The data was subjected to three treatments which were based on testing the data transformation. The method of clustering was kept constant and Ward’s Method was used. Ward’s Method minimizes the within cluster sum of squares at each agglomeration. Any difference in result will therefore be due to the differences in the dissimilarity matrix used.

- Fig5.3 and Fig.5.4 show the dendrogram and relative change in agglomeration coefficient for the data which was first Z-transformed. The squared Euclidean distance was used in the clustering algorithm which used Ward’s method for determining cluster agglomeration. Inspection of Fig.5.4 suggests a 3 or 7 cluster solution. Beale’s pseudo F was performed to see if the 7 cluster solution was better than the 3 cluster; \( F^* = 2.65 \), \( F_{\text{crit}} = 2.54 \) at 95% level of confidence, which suggests that the 7 cluster solution is only marginally better than the 3 cluster solution. The cut-off point is shown by the red line running across the dendrogram.
The dendrogram was formed from the standardized values of the 52x230 data matrix of word counts. The squared Euclidean distance was clustered using Ward's method.

Although the labels are difficult to see, because of the diagrammatic nature of the dendrogram, it can be seen how samples of blocks of text from Luke and Acts, for example, are clustered together.

<table>
<thead>
<tr>
<th>KEY</th>
<th>2Es</th>
<th>2 Esdras</th>
<th>Jn</th>
<th>John</th>
<th>Si</th>
<th>Similitudes</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Acts</td>
<td>Lk</td>
<td>Luke</td>
<td>So</td>
<td>Wisdom</td>
<td></td>
</tr>
<tr>
<td>Ex</td>
<td>Exodus</td>
<td>M</td>
<td>Mandates</td>
<td>V</td>
<td>Visions</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>Genesis</td>
<td>Ro</td>
<td>Romans</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IC</td>
<td>1 Corinthians</td>
<td>Rv</td>
<td>Revelation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IIC</td>
<td>2 Corinthians</td>
<td>Si</td>
<td>Sirach</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Relative change in the agglomeration can be seen to rise when going from 7 clusters to 6. The other change occurs in going from 3 clusters to 2. Trend lines are shown in red.

The cases are partitioned into the 7 clusters as follows:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>2Es</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ex</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>G</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>IC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>IIC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Jn</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Lk</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>M</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Ro</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Rv</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>S</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Si</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>So</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>V</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>
Cluster 4 contains the Pauline Corpus, cluster 4 contains the Lukan corpus and cluster 7 contains *The Shepherd* corpus. These are impressive results, but otherwise it can be noted that cluster 1 contains almost everything else!

- Fig.5.5 and Fig.5.6 are the result of using the square root(data)+square root(data+1) transformation of the raw data. Inspection of Fig.5.6 suggests either an 11 cluster solution or a 15 cluster solution. These were checked by Beale's pseudo-F statistic.

At the 95% level of confidence the 11 cluster solution is better than the 15 cluster solution.

**Fig.5.5** Dendrogram obtained from 60x230 matrix after values had been transformed by SQRT(data)+SQRT(data+1), followed by Ward's clustering method on the squared euclidean distance transformation.
Relative change in the agglomeration coefficient shows a slight increase in going from 16 to 15 clusters and 11 to 9 clusters. Trend line is shown in red.

The cases are partitioned into the 11 clusters as shown in the table below.
The Pauline corpus is contained in cluster 8. The Lukan corpus is contained in cluster 7. The Shepherd corpus is contained in cluster 10. The partition of the data seems promising. The SQRT transformation seems to be much better than the Z-transformation for discrimination.

- The proximity matrix from the Random Forest analysis of the previous chapter was next used. The agglomeration plot suggests cluster solutions 10, 9, 8, 7, 6 or 3. Beale’s pseudo-F suggests a 6 cluster solution.
Fig. 5.7 Dendrogram produced by the RF-proximity matrix clustered using Ward's method.
Relative change in the agglomeration coefficient suggests a clustering solution of 11, 10 or 7 clusters.

Beale's Pseudo F at 95% confidence favours a 7-cluster solution. The cases are partitioned as shown in the table below:
The Pauline corpus is found in cluster 5. *The Shepherd* is clustered in cluster 7. The Lukan corpus is found in cluster 4 together with Wisdom of Solomon.

Discussion of results will be deferred until non-hierarchical clustering has been performed. For the sake of completeness, Fig. 5.9 shows the dendrogram of the RF proximity matrix that has been constructed by divisive clustering.
Dendrogram of RF-Proximity Matrix By Divisive Clustering Using DIANA.

As with binary partition trees, divisive clustering tends to proceed by splitting the clusters by one variable at a time (monothetic). It is almost impossible to find the optimal divisions\(^94\), as the plot of the relative agglomeration coefficient shows in Fig.5.10.

\(^{94}\) Gordon, A.D. Classification. p.90
Divisive clustering tends to be used to discover a few of the large clusters only. Fig. 5.9, for example, would give confidence in asserting 4 clusters only.

Because of its limitations divisive clustering will not be considered as a viable method.

Nonhierarchical Clustering

This requires the setting of cluster 'seed points' around which clusters are built based on the distance between a case and a cluster seed. Clusters that are close may later be joined, but the process requires the researcher to guess the number of clusters before hand. The seed point can be set manually, that is at the discretion of the researcher, or, most packages will allow automatic choice by the computer. The problem with this latter arrangement is that usually this will be dependent on the order the data is input. The clustering will also be very dependent on the choice of seed. A poor initial choice will result in a poor clustering arrangement.
Although the procedure is criticized "because there are just too many possible choices, for not only the number of clusters, but for location of cluster seeds as well"95, there are ways around solving the problem. That is by use of the silhouette index96 introduced by Kaufman:

$$S(i) = \frac{(b(i) - a(i))}{\max\{a(i), b(i)\}}$$

where $a(i)$ is the average dissimilarity of i-object to all other objects in the same cluster as i; $b(i) = \text{minimum of average dissimilarity of i-object to all objects in other cluster (in the closest cluster)}$.

If $S(i) = 1$ then the sample is well clustered.

If $S(i) = 0$ the sample lies between two clusters by about an equal distance. It could therefore be assigned to either cluster.

If $S(i) = -1$ then the sample is "misclassified".

The average silhouette width for the entire plot is the average of the $S(i)$ for all cases in the whole data matrix. The optimal number of clusters is therefore the one that offers maximum $S(i)$. This was done for the three sets of data as above.

- Z-transformed, Euclidean distance matrix.

95 Johnson, D.E. *Applied Multivariate Methods* p.323
Silhouette Plot of Z-Transformed Data Using PAM with 4 Clusters

- Average Silhouette Width

\[ \text{Average silhouette width: 0.1} \]

- Cluster 1: 51 | 0.10
- Cluster 2: 5 | 0.15
- Cluster 3: 1 | 0.00
- Cluster 4: 5 | 0.07

\[ n = 52 \]

\[ 4 \text{ clusters } C_j \]

\[ j \text{ s.t. } \bar{s}_{\text{ave}} \]
These two components explain 16.92% of the point variability.

Fig. 5.11
(a) The plot of average silhouette width to determine optimal clustering. The largest silhouette width is for 4 clusters.
(b) Shows the Silhouette Width for the 4 cluster solution. Cluster 3 has a Sil.Widh of zero, therefore there are only three measurable clusters. Cluster 1 contains most of the cases, but the Sil.Width for each of the three clusters is small suggesting that the clusters are not very reliable.
(c) A two dimensional cluster plot of the data confirms the poor clustering of the data.
The ‘best’ clustering criteria is the one with the greatest value of the average silhouette width. The graph [Fig. 5.11(a)] shows that a 4 cluster solution is best. However, a plot of the silhouette width for each cluster [Fig. 5.11(b)] shows that the solution is not as good as would be hoped. This plot shows that only three clusters are significant, cluster 3 being zero. Positive values of silhouette width show good clustering, negative values suggest a poor, wrongly clustered case. The problems become evident when a clusterplot of the results is given [Fig. 5.11(c)], there is only small variability (16.9%) and clusters tend to overlap.

- SQRT-transformed data.
Silhouette Plot Of SQRT-transformed Data Showing 9 Clusters

$n = 62$

9 clusters $C_j$

$$j: \eta_j \text{ ave}_{\text{seg}} s_j$$

1: 4 | 0.39
2: 5 | 0.35
3: 3 | 0.27
4: 13 | 0.08
5: 14 | 0.12
6: 6 | 0.21
7: 6 | 0.25
8: 8 | 0.15
9: 3 | 0.36

Average silhouette width : 0.19

(b)
These two components explain 41.1% of the point variability.

(c)

**Fig. 5.12**

(a) Plot of silhouette width shows a maximum at 9 clusters.
(b) Sil.Width plot for 9 cluster solution. The larger values for Sil.Width suggests good partitioning of data to clusters.
(c) Two dimensional clusplot confirms good clustering with 41% of variance explained by the first two components.

- RF Proximity Matrix.
Average Silhouette Width Using RF-Proximity Matrix

Silhouette Plot of RF-Proximity Matrix For 14 Clusters

14 clusters $C_i$

$n = 62$

Average silhouette width: 0.1
These two components explain 16.72% of the point variability.

(a) Maximum Sil. Width is found at 14 clusters. All the Sil. Widths are small in value.
(b) Silhouette plot for 14 clusters. Cluster 4 is the worst defined of the clusters.
(c) Two dimensional plot of the 14 cluster solution.

Discussion

Divisive hierarchical clustering (DIANA) can be excluded immediately. The difficulties in interpretation of the dendrogram and the failure of the stopping rules to optimise the partitioning of the data are serious obstacles.

Hierarchical agglomeration was more successful. Note that only Ward’s clustering algorithm was used at this stage in order that comparisons of dendrograms from other data transformations could be made. This allows the structural efficiency
of the data transformation to be assessed. Since Ward’s method seeks to minimize the between-clusters sum of squares at each fusion, it produces results which are meaningful for interpretation. Also, the dendrogram that is produced is easy to interpret because it does not give any reversals that can complicate the picture (as some other clustering algorithms can do). As already stated on a previous occasion, results will always be produced by these methods consequently we need to carefully examine the results produced.

The Z-transformed data appears to have partitioned the data into 7 clusters, which can be compared with the other two transformation methods (SQRT-transformation giving an 11 cluster solution, the Random Forest proximity matrix giving a 7 cluster solution). The table of partition of cases to clusters also seems to give good results for the Z-transformation, that is, results that would suit the thesis that is being pursued! It also clusters The Shepherd corpus, the Pauline corpus and John’s Gospel; cluster 1 contains a mix of cases. The SQRT-transformation also performs very well. The Random Forest proximity matrix gives promising results similar to the Z-transformation.

The cophenetic correlation, that is a measure of how well the clustering structure relates to the original proximity matrix, can be easily calculated. Like the familiar correlation coefficient, the cophenetic correlation is a product moment correlation, in this case between the proximity matrix and the fusion heights of the dendrogram. It allows a comparison of the transformations to be made. The cophenetic correlations of the three proximity matrices and their respective dendrograms are:

1. Z-transformed cophenetic correlation = 0.546
2. SQRT-transformed cophenetic correlation = 0.645
3. RF-proximity cophenetic correlation = 0.605

The Z-transformed data actually performs slightly worse than the other two, and the results derived using it may not be as trustworthy as first assumed. The Z-transformation is not very good for clustering the data, the majority of the clustering being possible by pure chance.

Although the SQRT-transformed data gives a higher cophenetic correlation than the Z-transformed data, the eleven cluster solution is not convincing. However,
those clusters are slightly more reliable partitions of the data than those produced by the Z-transformation.

The RF-proximity matrix also performs reasonably well. The stopping rule did not work so well, but coupled with Beale's pseudo-F, a 7 cluster solution was found. The table of cases partitioned to the clusters also shows reasonable results, the Pauline corpus and *The Shepherd* corpus are extracted, but the Lukan corpus is clustered with Wisdom of Solomon. Genesis and Exodus show some interesting structure.

The cophenetic correlations are reasonably close to each other in value and we should not read too much into the differences.

The non-hierarchical method, PAM, gave good results. The Z-transformed data gave a poor 3 or 4 cluster solution. We can 'see' the quality of the clusters produced by examining the graphs of silhouette width. Experience suggests that silhouette widths under 0.2 are indicative of poorly formed clusters. We can see that although Z-transformed data can be clustered into 4 distinct clusters, the quality of those clusters is very poor indeed. They all have low silhouette indices, some even zero! This confirms the suspicion of the previous trial that the clusters are the product of chance and that only 2 clusters can be positively identified. Although Z-transformation *appears* to work well, the clusters it produces are not acceptable.

The SQRT-transformed data provided a 9 cluster solution. The silhouette plot shows good cluster performance with silhouette widths in the range of 0.4 (compared to the 0.06 of Z-transformation). There are some possible misclassifications in the solution, indicated by the negative silhouette values. The cluster plot of the first two components explain just under half of the total variability.

The RF-proximity matrix performed remarkably well, finding 14 clusters, but the silhouette plot shows that some of these clusters have very low silhouette indices. The cluster plot of the first two components only accounts for 16.9% of the total variability (similar to Z-transformed data). The best that this transformation has to offer is still half of what is given by the SQRT-transformation.
Conclusion

The best methodological options seem to be that of PAM clustering on the unsupervised Random Forest proximity matrix.

This chapter has confirmed the suspicion that Z-transformation of this type of data is NOT the best option. Its indiscriminate use could be resulting in erroneous conclusions being drawn from the data. One of the main criticisms of clustering procedures is the danger they pose for over-interpretation. This chapter has shown that this criticism would certainly apply if the ‘black-box’ approach was used on Z-transformed data.

The next chapter shows how this error could be compounded when used with Principal Components Analysis, yet PCA using the Z-transformed data appears to be the method that is mostly used in modern studies.
Chapter 6

Principal Component Analysis (PCA) and Hierarchical Clustering.

Introduction

Principal Components Analysis (PCA) has been used in many of the investigations of authorship by function word frequency count. Researchers have claimed that good results are obtained when this method is used. However, although graphs are presented which seem to confirm the discussions and conclusions of the researcher, much of the statistical information is not shown. This raises the suspicion as to whether the results are really as good as claimed, or has the researcher merely chosen the results that fit a pre-defined theory. Either case is a misuse of statistical procedure. It is easy to accept results at face-value, especially if the first run produces results that the researcher would like to see. The previous chapter has already shown that a Z-transformation and hierarchical clustering gave excellent results that would have confirmed research of this thesis immediately. This misuse of statistical inference is often exacerbated by the ease with which computer software allows statistical calculations to be performed. The ‘black box’ approach enables data to be entered into the computer and it then prints the results. Results will always be given, but they will not always be meaningful, especially if the data or statistical technique used is not valid. In this chapter a PCA of the data obtained (as outlined previously) will be evaluated to see how well PCA performs.

What is PCA?

Any textbook about multivariate statistics will contain a chapter on PCA, but for our purposes we can regard it as a method that partitions the total variance of the data into coherent subsets. Variables in the subset are correlated with one another, but are largely independent of other subsets of variables. If we start with a large number of variables, \( p \), PCA finds linear combinations of variables where the variance of those variables will be largest. By the combining of variables, the total variance that was present in the \( p \) variables is now partitioned into the \( g \) groups. Usually \( g \) will be less than \( p \). This is equivalent to reducing the number of data variables.
In order to achieve this reduction, the original matrix of data will need to be converted into the covariance matrix, or if preferred the correlation matrix, in order to find those variables which have the greatest covariance (or correlation) between them. The covariance matrix will be a square matrix whose number of columns is the same as the number of rows, that is a p by p matrix. Note that in doing this the labels on the original data (in our case the text from which the variables are measured) are NOT included in the matrix, that is, PCA is an unsupervised method. Matrix algebra shows that a square matrix is actually formed by the product of two, more basic matrices, namely the eigenvalue matrix and the eigenvector matrix. The eigenvalue matrix is a square matrix with the eigenvalues in the diagonal and zeros everywhere else. Thus there are p eigenvalues. The eigenvectors are a vector of values such that when multiplied by the eigenvalue matrix they will reproduce the data matrix. Unfortunately the eigenvectors are not unique in that any multiple of the eigenvectors would solve the equation. However, by normalizing (set that the sum of the squares of the eigenvectors = 1), the eigenvectors will have unique solution.

The sum of the diagonals of the eigenvalues matrix is equal to the total variance in the original data, but, the eigenvalues are arranged in such a way that the first eigenvalue represents the combination of variables with the greatest variance. The eigenvalues are therefore a measure of the partition of the variance between combinations of variables starting with the greatest and then decreasing in order. The eigenvalues should be reported, or at the very least a plot of the eigenvalues (known as a scree plot). This is very often not the case\(^7\).

When the normalized eigenvector matrix and the original data matrix are now multiplied together, the result is the unstandardized principal component score. This provides us with the labelled data again, so that now each case has its principal component score instead of its variable count. It is usual to find the first two principal component scores plotted against one another and the data points labelled in order to show the cases that tend to be closest to one another, or most similar in terms of their first two principal components.

The results of PCA are not so good if the number of variables is close to the number of cases. A ‘rule of thumb’ suggests that there should be a ratio of approximately 3 to 5 cases for every variable, and that the more cases there are, the

\(^7\) see for example Binongo, "Who wrote the 15th Book of Oz". He does tell us, however, that the 1st PC accounts for 20% of the total variance and the 2nd PC accounts for 7% of the variance. He finds that only 27% of the total variance is needed to distinguish between two authors.
better the method performs.

There are two main problems that need to be addressed as matters of great concern when using PCA, namely transformation and dimensionality. The first problem arises before we use PCA, a problem that we have already noted in the previous chapters, that of data transformation. The second is the problem of dimensionality.

**Dimensionality**

PCA is sometimes referred to as a dimension reduction technique. Starting with \( p \) variables, these are reduced to \( g \) components, where \( g < p \), and \( g \) is some linear combination of variable of \( p \). The components satisfying some criterion are retained. Usually this criterion is ‘keep the number of components whose eigenvalues are greater than 1’, or, ‘keep the number of components that account for a certain proportion of the variance e.g. 76%’. But, it must be remembered that PCA assumes that the sample of individuals is homogeneous. If the sample is homogeneous, then the first few components will be useful, since most of the variance will be explained by them. However, if there is heterogeneity between individuals i.e. a group structure to the data, then the usefulness of a component depends upon the extent to which the variance is related to the between-group variance. Bibby\(^98\) defines the distance between groups on component \( i \) as:

\[
\frac{(C_i - d_i)^2}{\lambda_i}
\]

and there is no guarantee that components with the largest eigenvalues represent the information that helps define the between-group variance. If the reduced PCA components are then used in further statistical analysis, such as cluster analysis or discriminant analysis, then the results may be misleading because the components which give the most structure information may have been discarded because they are less than the acceptance criteria. More importantly, all of those papers which plot only the first two principal components may be basing conclusions on misleading information!

**Data Selection**

In the previous two chapters the fact that the data has more variables than cases

---

\(^98\) Mardia,K.V., Kent,J.T., Bibby,J.M. *Multivariate Analysis.*
has not mattered, but now we need to have approximately 3 cases for each variable, and less would be better. How do we choose the variables to include? One method might be to take all of those variables whose mean is greater than a set value. This is not entirely satisfactory because, as we have seen, low count words can be important discriminators of structure. A better method might be to try to obtain the importance of each variable in its ability to define structure in the data. A method to do this will be shown later, but for now we will base selection on the variable being greater than a set value.

Results

Using a sample size of 2000 words results in 96 cases. Variables were chosen on the basis that mean count was greater than 1 which then resulted in a 96 x 41 working data matrix. The words used in the analysis are:

<table>
<thead>
<tr>
<th>καί</th>
<th>δέ</th>
<th>οὐ, οὐχ, οὐκ</th>
<th>ἐπί, ἐπ’, ἐφ’</th>
<th>μῆ</th>
</tr>
</thead>
<tbody>
<tr>
<td>δτι</td>
<td>ἐκ, ἐξ</td>
<td>γάρ</td>
<td>ἀπό, ἀπ’, ἀφ’</td>
<td>πρός</td>
</tr>
<tr>
<td>μετά, μετ’, μεθ’</td>
<td>διά, δι’</td>
<td>ἰνα, ἰν’</td>
<td>ὡς</td>
<td>ἀλλά, ἀλλ’</td>
</tr>
<tr>
<td>ἐάν, ἐάνπερ</td>
<td>κατά, κατ’, καθ’</td>
<td>ei, ei’peri, ei’pos</td>
<td>peri</td>
<td>ἦ, ἦτοι</td>
</tr>
<tr>
<td>παρά, παρ’</td>
<td>οὔτω, οὔτος</td>
<td>ὑπό, ὑπ’, ὑφ’</td>
<td>νῦν, νυνί</td>
<td>πάντως</td>
</tr>
<tr>
<td>μὲν</td>
<td>έκεῖ, έκεῖθεν</td>
<td>ἄν</td>
<td>ἥσσ’</td>
<td>ἐτι</td>
</tr>
<tr>
<td>οὐδὲ</td>
<td>καθὼς</td>
<td>ὑπέρ</td>
<td>ὅταν</td>
<td>πάλιν</td>
</tr>
<tr>
<td>ἐνώπιον</td>
<td>οὔτε</td>
<td>πῶς</td>
<td>ἀνά</td>
<td>σύν</td>
</tr>
<tr>
<td>ὦστε</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The data was Z-transformed and PCA performed.

Eigenvalues and % variance explained by them is shown in the table below. Only the first 15 eigenvalues are shown, but using the rule that eigenvalues greater than 1.00 are significant shows that the first 11 components are important and account for 70% of the variance.
The associated scree plot is shown for completion.

**Scree Plot Z-Transformed Data**

![Scree Plot Z-Transformed Data](image)

Fig. 6.1 Scree Plot of eigenvalues against the number of components from PCA. Components with eigenvalues less than 1 are not significant. In this case the first 11 components are significant.
The table of loadings of variables to components is shown Appendix 6.1.

The plot of the first two principal components shows some interesting structure, there is a clear separation of a group containing Romans, 1 Corinthians, 2 Corinthians and Galatians (the Pauline corpus). Sirach (Ecclesiasticus), Si, also forms a distinct group. It then gets a little more difficult to distinguish other groups.

**Principal Component Scores**

![Plot of first two principal components showing that there is some structure to the data. The first two component scores account for only 30% of the variance](image)

*Fig. 6.2*

*Plot of first two principal components showing that there is some structure to the data. The first two component scores account for only 30% of the variance*
Using the 11 Principal Components a hierarchical clustering method (Ward’s) can be used and the dendrogram plotted (Fig.6.3). As explained in the previous chapter, the optimal number of clusters is determined and the partition of cases to clusters tabulated.

**Dendrogram (Ward's Method) of PC[1:11]**

![Dendrogram of the first 11 Principal Components by Ward's Method. It shows a clear clustering of the Pauline corpus.](image_url)

The change in agglomeration plot (Fig.6.4) suggests 15, 8, 7, 6, 5 or 4 clusters. Beale's pseudo-F supported a 7 cluster solution.
Relative changes in the agglomeration coefficient can be noted in going from 15 to 14 clusters, 8 to 7 clusters and from 5 to 4 clusters.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>2Es</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ex</td>
<td>4</td>
<td>0</td>
<td>7</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>G</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Ga</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>IC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>IIC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Jn</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Lk</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>M</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ro</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Rv</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>S</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Si</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>So</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>V</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Discussion

The table of cases partitioned to clusters gives better reading than the plot of the first two principal components. Some clusters are still not easy to explain. 2 Esdras, Exodus, Genesis, Revelation and part of Ecclesiasticus are together in the same cluster (cluster 1) but cluster 2 separates John’s Gospel. Cluster 3 has all of Luke and Acts together with most of Exodus and doesn’t show any structure in these texts. Cluster 4 contains The Shepherd corpus along with The Book of Wisdom (a pattern which seems to be fairly consistent), but it also contains one section from Exodus. Cluster 5 has separated the Pauline corpus, cluster 6 most of Ecclesiasticus. Cluster 7 has most of Genesis.

Although the text separation is not that good, it probably has to do with the fact that there is a poor ratio of cases to variables. Also, we have used the top count variables, which may not be the best to use in order to detect structure. However, because variables with highest frequencies are used there is a possibility of using smaller sample sizes\(^9\). The main disadvantage is that there is no scientific reason for choosing variables, apart from the fact that using those whose mean was greater than 1.0 just happened to give enough variables in order to perform a PCA. This is not a very satisfactory situation, and the choice of variables could not be defended on any logical grounds.

All of this suggests that PCA should NOT be the primary statistical tool used in the analysis, as most researchers have tended to assume, but rather it is an analysis used to confirm or supplement other preliminary methods. This is because PCA is arbitrary, in the sense that different results will be obtained if different variables are included in the analysis. Deciding which variables to include may also be arbitrary, unless a clear method of selection is used. The variability of the method is seen also in the fact that the same variables when transformed by different transformation algorithms eg. Z, square root or log, will produce different results. Therefore a better method is required and the next few chapters will deal with some of the other statistical methods that might be used in this quest.

---

Chapter 7

OPTIMAL VARIABLE WEIGHTING (OVW)

AND

MINIMUM E-DISTANCE CLUSTERING

Introduction

This chapter assesses two other methods. The first method is Optimal Variable Weighting, and its developers claim good results from its use.

The second method is Hierarchical Clustering by Minimum (Energy) E-distance. The developers claim it is superior to other, more traditional, methods of hierarchical clustering.

Both of these are tested on the original data (64x236) matrix.

Optimal Variable Weighting (OVW)

Makarenkov and Legendre\textsuperscript{100} use the Taguchi loss function that was further developed for tree-based statistical methods by De Soete\textsuperscript{101}. The loss function was originally introduced to minimize financial loss in economic forecast models. Given a polynomial whose equation is given by:

$$d_{ij} = \left( \sum_{k=1}^{p} w_k (x_{ik} - x_{jk})^2 \right)^{1/2}$$

where $w_k$ is a weighting factor, $p$ is the number of variables.

Then, $d_{ij}$ can be predicted, based on values (unknown) of $w_k$. The predicted $d_{ij}$ will differ from the actual value (which is known) because of wrong choices in $w_k$, the amount of difference is the forecast error $e_k = (d_{ij} - d_{ij}^{\text{pred}})$. The expected value of the squared forecast error is therefore:

\textsuperscript{100} Makarenkov, V. and Legendre, P. "Optimal Variable Weighting for Ultrametric and Additive Trees and K-means Partitioning: Methods and Software". \textit{Journal of Classification} (2001) 18 pp.245-271

\textsuperscript{101} De Soete,G. "OVWTRE: A program for Optimal Variable Weighting for Ultrametric and Additive Tree Fitting". \textit{Journal of Classification} (1988) 5 pp.101-104
\[
E(e_k^2) = E \left[ f((x_{ik} - x_{jk})w_k) + e_k - d_{ij}^{pred} \right]^2
\]

This is minimised when

\[
f((x_{ik} - x_{jk})w_k)^2 = d_{ij}^{pred^2}
\]

The values of \( w_k \) can be chosen to minimise the expected error. In order to do this, since \( w_k \) is not known, initial guesses of \( w_k \) have to be made and better values calculated. These values then act as the guesses in another round of calculation and even better values calculated. A mathematical function is required to do this iterative process until the values of \( w_k \) no longer change. This optimisation of \( w_k \) is achieved by using the method of conjugate gradient optimisation of Polak-Ribiere and was chosen by Makarenkov and Legendre\(^{102}\) simply because it is the quickest method (converges rapidly) when working with a large data matrix. A process of iteration finds the optimum values of \( w_k \). Because there may be a large number of solutions, the normalised weights are reported, that is, the weights \( w_k \) are subject to the conditions:

1. All individual weights are greater than, or equal to, 0.
2. The sum of weights = 1

Makarenkov and Legendre claim that this method enables the best clustering results. However, the method down-weights those variables that have the highest values. In some instances, these high value variables can be given zero weight if tolerances are not carefully adjusted!

**Results**

Using the original 64x256 matrix of raw data, the vector of optimal weights \( W \) found for the variables is given in Appendix 7.1. Most of the variable weights are small because their sum must only be equal to 1. A dissimilarity matrix based on the optimal weights is also given as part of the program. This dissimilarity matrix can then be used in hierarchical clustering methods, and by using Ward’s method again,

\(^{102}\) A freeware computer program, OVW, is available from Makarenkov at makarenv@ere.umontreal.ca
we are able to compare how the OVW performs in comparison with Z-transformation, SQRT-transformation and RF-proximity matrix.

**Fig1.1a** The relative change in the agglomeration coefficient suggests either 5, 6, 9, 10 or 11 clusters.

*Beale's Pseudo-F suggests a 5 cluster solution.*
The results indicate a 5 cluster solution is found, which partitions the cases into the following clusters:
The cases are partitioned in a random manner suggesting that the discriminatory power of this method is not much better than random chance. Examination of the coefficients in Appendix 7.1 shows that the method heavily penalizes large count data (their coefficients tend to zero!).

We can conclude that using all of the variables in a weighted matrix does NOT give sufficient discriminatory power to warrant its further use.

E-Distance Clustering

The data can also be analysed using a multivariate measure of the distance between distributions. All of the above methods measure distance between points and are based on the sum of the differences between points over all variables. This does not take into account the distribution of results for each of the variables measured, since the distribution is unknown. However, it is possible to get a reliable estimate of the empirical distribution function (EDF) using bootstrap methods. Székely and Rizzo used this approach in their study of high dimensional data.

103 Davison A.C., Hinkley D.V. *Bootstrap Methods and Their Applications*. This provides a detailed examination of bootstrap methods, with numerous practical examples.
They used the Euclidean distance as their starting point, but define the distance between two distributions, \( A = \{a_1, a_2, \ldots, a_n\} \) and \( B = \{b_1, b_2, \ldots, b_n\} \) as the e-distance, which, for \( k=2 \) is defined by:

\[
\epsilon(A, B) = \frac{n_1 n_2}{n_1 + n_2} \left( \frac{2}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \|a_i - b_j\| - \frac{1}{n_1} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \|a_i - a_j\| - \frac{1}{n_2} \sum_{i=1}^{n_2} \sum_{j=1}^{n_1} \|b_i - b_j\| \right)
\]

where \( \| . \| \) is the Euclidean distance.

The e-distance is therefore a measure of both the heterogeneity between clusters and the homogeneity within clusters.

Using the three transformations of the previous chapters, based upon the 62x230 matrix allows direct comparison between the techniques "Ward" and "E-distance".

**Results**

- **Z-Transformed Data.**

![Change in Agglomeration Coefficient](image)

*Fig.7.2* Change in agglomeration coefficient for Z-transformed e-distance clustering. The first significant increase in coefficient occurs going from 9 to 8 clusters. A small increase is noted going from 6 to 5 clusters. A large increase occurs going from 3 to 2 clusters. Trend line shown in red.
Cophenetic Correlation = 0.329

- SQRT-Transformed Data.
Cluster Dendrogram based on e-distance of \( \sqrt{\text{data}} + \sqrt{\text{data}+1} \) transformation.

Cophenetic Correlation = 0.685

- RF-Proximity Matrix.

Change in Agglomeration Coefficient

Change in agglomeration coefficient for RF-proximity matrix by e-distance clustering. Cluster solutions between 13, 10, 9, 8, 7 are possible.
Cophenetic Correlation = 0.622

Discussion

Comparing the cophenetic correlation between “Ward’s method” and “E-distance”, it can be seen that the trend is Z-transformation < RF-proximity < SQRT-transformation. The “E-Distance” does perform better than “Ward’s method”.

Conclusion

In this chapter we have been able to confirm again the inadvisability of the Z-transformation, which will now be dropped from all further consideration. Optimal Variable Weighting was also a poor performer and will be dropped from further consideration. In view of the good performance of the E-distance clustering method, we will take it a step further in the next chapter and consider density mixtures.
Consider the word $\delta t \dot{\alpha}$ found in the 3000 word sample size blocks of texts of interest. The plot of a histogram of the frequency of counts is shown, together with the density estimation of the same data using a band width calculated from ‘Sheather-Jones direct plug-in’.

**Frequency Histogram an Density Estimate**

![Frequency Histogram and Density Estimate](image)

*Fig. 8.1* The frequency histogram and density estimate (black line) using the Sheather-Jones direct plug-in kernel for word-count taken from the 3000 word sample block of text. It can be seen that the count is discontinuous and complex. It is not a single 'bell-shaped' curve, but is a skewed density with up to 6 peaks.
The normal (Gaussian) density, the famous ‘bell-shaped’ curve is not found. Hence we can see part of the failure of the $Z$-transformation, which assumes a normal curve in calculating mean and standard deviation. Inspection of our curve seems to suggest that it is the result of the mixture of several normal Gaussian kernels. This can be represented pictorially:

![Pictorial representation of mixed density estimate for mixture modelling. The single density curve can be thought as deriving its shape from the six 'normal' constituents being added together.](image)

**Fig. 8.2** Pictorial representation of mixed density estimate for mixture modelling. The single density curve can be thought as deriving its shape from the six 'normal' constituents being added together.
The original data therefore consists of a mixture of normal Gaussian kernels. The finite mixture model approximates the unknown density $p(x)$ by a finite weighted sum of normal Gaussian densities, expressed as:

$$p(x) = \sum_{j=1}^{K} \pi_j f_j(x)$$

where $\pi_j$ is the mixing weight, $K$ is the number of kernels, and $f_j(x)$ is the univariate normal density with mean $\mu_j$ and standard deviation $\sigma_j$, ie.

$$f_j(x) = \frac{1}{\sigma_j \sqrt{2\pi}} \exp \left[ -\frac{(x - \mu_j)^2}{2\sigma_j^2} \right]$$

Using the data, the vector of $x$ variables, we have to calculate the vector of the unknowns $\Psi = [\mu_1, \sigma_1, \mu_2, \ldots, \mu_K, \sigma_K]$. $\Psi$ is called the likelihood function and by process of iteration the values which maximise the logarithm of the likelihood function are found. This gives the Maximum Likelihood (ML) estimate of $\Psi$.

We can generalise the above to the multivariate case where matrices are employed:

$$f_k(x|\mu_k, \Sigma_k) = \frac{1}{\sqrt{(2\pi)^d \Sigma_k}} \exp \left[ -\frac{1}{2} (x - \mu_k)^T \Sigma_k^{-1} (x - \mu_k) \right]$$

Where $x$ is now the matrix of variables, $\mu$ is a matrix of mean values, $\Sigma$ is the covariance matrix, $k$ is a cluster and $d$ is the dimension of the data.

Each covariance matrix can be parameterized by its eigenvalue decomposition, ie.

$$\Sigma_k = \lambda_k D_k A_k D_k^T$$

Where $D_k$ is the orthogonal matrix of eigenvectors, $A_k$ is a diagonal matrix whose elements are proportional to the eigenvalues of the covariance matrix and $\lambda_k$ is a scalar multiplier. Fraley and Raftery\textsuperscript{105} show that the shape and volume of the cluster

---


ellipsoid is governed by the values of the decomposed covariance matrix. The cluster shape can be spherical, diagonal or ellipsoidal and various combinations of either equal or variable volume and equal or variable shape.

The SVD reduced matrix estimation of finite mixture density clustering can be found in R EMclust.

Since EMclust involves matrix inversion, the method is subject to problems of matrix multicollinearity and ill-conditioned matrices.

Results

- SQRT-transformed data.
Fig. 4.3  EMclust of SQRT-transformed data using the same data matrix as used for PCA in chapter 6. The best model is 8 clusters type D (diagonal, varying volume, equal shape), although there is not a great deal of difference between this model and models C, B or A.

BIC (y-axis) is the Bayesian Information Criterion.
We can see that 2Esdras is in a cluster of its own (cluster 1), as is John’s Gospel (cluster 2). Genesis and Exodus are split over three clusters with clusters 3 and 4 having parts of Genesis and Exodus in common. This indicates a complex editorial process has been involved in the construction of these two texts. Cluster 5 contains *The Shepherd* corpus, but includes The Book of Wisdom (again) and a single case from Exodus. The Pauline corpus is grouped together into a single cluster of its own. The Lukan corpus is clustered together, but contains one case from Genesis. Revelation and Ecclesiasticus are included together in cluster 8.

- RF-proximity matrix.

This method does not work well when using the proximity matrix from unsupervised Random Forest.

**Discussion**

The finite density mixture model works extremely well on data where the number of variable is less than the number of cases. Since the same data is used in this analysis that was used for PCA analysis in chapter 6, we can compare the results
obtained. It becomes clear that SQRT-transformed finite density mixture clustering is far superior to Z-transformed PCA. This result was expected due to the skewed, non-continuous variable counts that do not give favourable transformations with skewed, non-continuous variable counts.

Although finite density mixture clustering does not give good results when the RF proximity matrix of unsupervised Random Forest is used, it is still a good method to use on reduced data. Since using only the higher frequency words allows smaller sample sizes to be chosen, finite density mixture clustering would be a useful method to use to supplement the results obtained from unsupervised Random Forest and PAM clustering methods.

Conclusion

Finite mixture density works extremely well with the data obtained from word counts in texts. It uses a better working model for determining clusters, being based on the frequency distribution of the word over all the texts. The method could be refined if the Kullback-Liebler function was used to compare frequency distributions, a mathematical refinement that is beyond the scope of this thesis, but should be evaluated at a later date.

Having evaluated a number of statistical methods, the next chapter describes which methods will be used in a working procedure, and will test their sensitivity and robustness.
Chapter 9

Working Procedure

Testing and Refining

Working Procedure Proposed

Based on the results of the previous chapters, a basic working procedure can now be outlined.

1. The chosen texts are reduced to individual consecutive word blocks of given size. The number of function words in each block is determined. A fixed master-set of function words is used. The fixed master-set contains 303 function words. (See Appendix 9.1).

2. **Unsupervised Random Forest** is performed on the resulting data matrix of word counts and the proximity matrix obtained.

3. **PAM clustering** is used on the proximity matrix from step 2. and the optimal number of clusters obtained by maximizing the silhouette width. A table of cases partitioned amongst clusters is obtained [Results Table 1].

4. **A supervised Random Forest** is conducted on the data matrix from step 1 and the cross validation classification table obtained [Results Table 2]. Variable importance is used to produce a subset of data based on the important discriminatory words only, ie. a reduced data matrix.

5. **PAM clustering** is used on the proximity matrix from step 4 and the optimal number of clusters obtained by maximizing the silhouette width. A table of cases partitioned amongst clusters is obtained [Results Table 3].

6. **FMD** is used on the sub-set data matrix from step 4 and the tables of cases partitioned amongst clusters obtained [Results Table 4].

7. The results from the 4 tables are compared and conclusions drawn.

Working Procedure Tested.

Using this working procedure the methodology can be refined.

a) Sample Size
The sample size has consistently been held at the 3000 word block size. This value was chosen because it fitted the sizes of the texts used, the shortest text containing just over 4400 words. This sample size also allowed two blocks of text to be drawn from Visions and Mandates (both just greater than 6000 words in length). The method is now tested, following the working conditions set out above, in order to discover at what point the sample size variances fail to distinguish between cases. This is done by using the same texts and variables that have been used in the past chapters, but now the sample sizes are chosen as 3000, 2500, 2000, 1500, 1000 and 500 word blocks of consecutive text.

Using unsupervised random forest and PAM clustering, the maximum silhouette width is used to obtain the optimum number of clusters. This was done for each sample size. The number of clusters found are shown in the table below. A clear break-down of the ratio of within-cluster sum squares and between-cluster sum squares can be noted from the cluster solutions suggested, which are:

<table>
<thead>
<tr>
<th>Block Size</th>
<th>Clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td>3000</td>
<td>14</td>
</tr>
<tr>
<td>2500</td>
<td>11</td>
</tr>
<tr>
<td>2000</td>
<td>10</td>
</tr>
<tr>
<td>1500</td>
<td>14</td>
</tr>
<tr>
<td>1000</td>
<td>2</td>
</tr>
<tr>
<td>500</td>
<td>2</td>
</tr>
</tbody>
</table>

Structural information within the data is lost when the sample size is less than 1500 words (only 2 clusters found). Sample sizes greater than 1500 pick-up the more detailed structure within the data, and seem to vary around an average 11 to 12 clusters. This can be clearly seen in the tables of partitions of cases into clusters.
### Sample Size = 500

<table>
<thead>
<tr>
<th>Cluster</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>2Es</td>
<td>26</td>
<td>0</td>
</tr>
<tr>
<td>A</td>
<td>3</td>
<td>32</td>
</tr>
<tr>
<td>Ex</td>
<td>27</td>
<td>17</td>
</tr>
<tr>
<td>G</td>
<td>27</td>
<td>38</td>
</tr>
<tr>
<td>IC</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>IIC</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Jn</td>
<td>4</td>
<td>27</td>
</tr>
<tr>
<td>Lk</td>
<td>6</td>
<td>32</td>
</tr>
<tr>
<td>M</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>Ro</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>Rv</td>
<td>19</td>
<td>0</td>
</tr>
<tr>
<td>S</td>
<td>1</td>
<td>28</td>
</tr>
<tr>
<td>Si</td>
<td>3</td>
<td>16</td>
</tr>
<tr>
<td>So</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>V</td>
<td>0</td>
<td>12</td>
</tr>
</tbody>
</table>

### Sample Size = 1000

<table>
<thead>
<tr>
<th>Cluster</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>2Es</td>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>A</td>
<td>2</td>
<td>15</td>
</tr>
<tr>
<td>Ex</td>
<td>21</td>
<td>3</td>
</tr>
<tr>
<td>G</td>
<td>26</td>
<td>6</td>
</tr>
<tr>
<td>IC</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>IIC</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Jn</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>Lk</td>
<td>3</td>
<td>16</td>
</tr>
<tr>
<td>M</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Ro</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Rv</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>S</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td>Si</td>
<td>5</td>
<td>13</td>
</tr>
<tr>
<td>So</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>V</td>
<td>0</td>
<td>6</td>
</tr>
</tbody>
</table>

### Sample Size = 1500

<table>
<thead>
<tr>
<th>Cluster</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>2Es</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ex</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>9</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>G</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>IC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>IIC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Jn</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Lk</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>M</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ro</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Rv</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>S</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Si</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>So</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>V</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Sample Size = 2000

<table>
<thead>
<tr>
<th>Cluster</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>2Es</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ex</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>6</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>G</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>12</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>IC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>IIC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Jn</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Lk</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>M</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ro</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Rv</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>S</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>Si</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>So</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>V</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
</tbody>
</table>

Sample Size = 2500

<table>
<thead>
<tr>
<th>Cluster</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>2Es</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ex</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>G</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>9</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>IC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>IIC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Jn</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Lk</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>M</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ro</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Rv</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>S</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Si</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>So</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>V</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
The increase in the number of clusters in the larger sample sizes resolves Luke-Arts into clusters which tend to be separated from any of Genesis-Exodus, and splits the Mandates and part of Similitudes from the rest of *The shepherd*.

From this we conclude that a minimum Sample Size would consist of 1500 words (most researchers have used 1000 word blocks as standard practice and have not questioned the adequacy of the sample size). In all that follows a Sample Size of 2000 has been used.

**Additional Texts**

The robustness of the working procedure was checked by repeating the process (using the same function words for counts), but adding another text each time. The tables of partitions of cases to clusters by PAM are shown below. The texts added were:

1. Galatians
2. Isaiah
3. 4 Maccabees
4. Jeremiah
5. Matthew
1. Addition of Galatians:

<table>
<thead>
<tr>
<th>Cluster</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>2Es</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ex</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>6</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>G</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>12</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ga</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>IC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>IIC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Jn</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Lk</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>M</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ro</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Rv</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>S</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Si</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>So</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>V</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Galatians had been included in the texts for PCA and Finite Mixture Density analysis because these methods allowed smaller sample sizes, it is now included in the unsupervised Random Forest/PAM analysis. It is one of the undisputed letters of Paul and it would be expected to be included in the Pauline Corpus cluster. Examination of the table above shows that this is the case.

2. Addition of Isaiah:

Isaiah was added as a check on the ability of the method to distinguish it from Genesis/Exodus as major Old Testament texts.
3. Addition of 4 Maccabees:

4 Maccabees was added to balance the Hebrew/Greek text ratio.
4. Addition of Jeremiah:

Jeremiah was added to check the ability of the method to distinguish prophetic genres.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>2Es</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4M</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>Ex</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>G</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>12</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ga</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>IC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>IIC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Is</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>12</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Jn</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Jr</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Lk</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>M</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Ro</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Rv</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>S</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>Si</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>So</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
</tbody>
</table>

5. Addition of Matthew:

Matthew was added as the biggest test, to check Gospel genre with Luke, and, whether any Luke-Matthew structure could be found.
Discussion

It can be noted from the tables that the initial stages of the working method are robust enough to handle the extra data. The final table (above) shows some interesting clustering. Exodus is split over four clusters, Genesis over two. As previously suggested, this indicates a number of editorial changes to the texts. Matthew and Luke have a number of cases in the same cluster (cluster 5). This is expected due to the common sources used by these two authors. It is surprising to find the whole of Matthew in the cluster with some of Luke, but this may be due to the choice of variables (remember, these variables have been the ones chosen for the initial texts and have not been up-dated with the addition of new texts).

The Pauline corpus has remained a consistent cluster throughout, as has The Shepherd corpus (apart from a dislocation of one case from the Visions in the final table).

One of the greatest surprises is the fact that all cases of Isaiah were clustered in one cluster.
Working Method Sensitivity

Returning to the original texts (before the additions of texts to test robustness), the sensitivity of the method to foreign material was assessed. This was achieved by contaminating one section of Similitudes with text from Revelation. A Random Forest analysis, followed by PAM clustering of the proximity matrix, gave a table of clustered groups. By varying the amount of contamination in the sample, the sensitivity of the method to detect the change can be obtained. The table of contamination % and result is shown below:

<table>
<thead>
<tr>
<th>Similitude contaminated with Revelation</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>30%</td>
<td>Contaminated Similitude appears in its own group with 1 Similitude sample</td>
</tr>
<tr>
<td>20%</td>
<td>Contaminated Similitude appears in its own group with 1 Similitude sample</td>
</tr>
<tr>
<td>10%</td>
<td>Contaminated Similitude appears in its own group with 1 Similitude sample</td>
</tr>
<tr>
<td>5%</td>
<td>Contaminated Similitude appears in its own group with 1 Similitude sample</td>
</tr>
<tr>
<td>2%</td>
<td>Contaminated Similitude appears as a singlet in its own group.</td>
</tr>
</tbody>
</table>

The method is sensitive enough to detect a block of words from Revelation at less than 2% of contamination. On a sample size of 2000 words, the method will detect a 40 word block inserted into the sample. A small proportion of those 40 words will be function words, so it can be seen that the method is sensitive. When the contamination is 5% or greater, the method separates the contaminated sample from the original uncontaminated cluster, but since the original uncontaminated sample was retained in the analysis, it tends to draw the original from its cluster into the contaminated group, forming a cluster of two samples. We would not have this luxury in samples taken from texts, but what these results show is that only a small amount of text from another author will split the samples across different clusters. Therefore, samples of a text split across several clusters will be indicative of authorship impurity.

106 Note the limitations of this test. Only one section of Similitudes is used and only one type of contaminant text. Therefore, the sensitivity obtained here is a guide only. To get an analytical result would require testing on many more sections of text and a variety of contaminant texts of different genres. This is beyond the scope of this thesis, nevertheless, this test is a guide as to methodology limitations.
RESULTS

We can now use the proposed working method in order to determine whether *The Shepherd* is the work of a single author or multiple authors or editors. Using a sample size of 2000 words, the following texts were included in the analysis:

<table>
<thead>
<tr>
<th>Ref</th>
<th>Text</th>
<th>Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>2Es</td>
<td>2 Esdras</td>
<td>6</td>
</tr>
<tr>
<td>4M</td>
<td>4 Maccabees</td>
<td>3</td>
</tr>
<tr>
<td>A</td>
<td>Acts</td>
<td>8</td>
</tr>
<tr>
<td>Ex</td>
<td>Exodus</td>
<td>12</td>
</tr>
<tr>
<td>G</td>
<td>Genesis</td>
<td>16</td>
</tr>
<tr>
<td>Ga</td>
<td>Galatians</td>
<td>1</td>
</tr>
<tr>
<td>IC</td>
<td>1 Corinthians</td>
<td>3</td>
</tr>
<tr>
<td>IIC</td>
<td>2 Corinthians</td>
<td>2</td>
</tr>
<tr>
<td>Jn</td>
<td>Gospel of John</td>
<td>7</td>
</tr>
<tr>
<td>Lk</td>
<td>Gospel of Luke</td>
<td>9</td>
</tr>
<tr>
<td>M</td>
<td>Mandates</td>
<td>3</td>
</tr>
<tr>
<td>Ro</td>
<td>Romans</td>
<td>3</td>
</tr>
<tr>
<td>Rv</td>
<td>Revelation</td>
<td>4</td>
</tr>
<tr>
<td>S</td>
<td>Similitudes</td>
<td>7</td>
</tr>
<tr>
<td>Si</td>
<td>Sirach (Ecclesiasticus)</td>
<td>9</td>
</tr>
<tr>
<td>So</td>
<td>Wisdom of Solomon</td>
<td>3</td>
</tr>
<tr>
<td>V</td>
<td>Visions</td>
<td>3</td>
</tr>
</tbody>
</table>

The samples column indicates how many consecutive 2000 word samples were taken from each text (99 cases in total).

There were 303 variables counted on each sample, 230 variables were non-zero. The words chosen for the analysis were:

\[\text{άχρι, άχρις, άφω, άλλως, άν, άνευ, άνω, άνωθεν, άρτι, άτερ, άδιαλείπτως, άδίκως, άεί, άφθος, άφρονός, άγνως, άκριβῶς, άληθῶς, άλλαξθεν, άλλα, άλλ', άμεμπτως, άναντι, άνδρείως, άνάμως, άντι, άνθ', άντ', άπέναντι, άπλως, άποτόμως, άπο, άπ', άφ', άσφαλως, αύριον, άμα, άπαξ, χαμαί, χαμίν, χωρίς, δεινόν, δεινό, δεώ, διή, διά, δι', δικαίως, διό, διόπερ, διότι, δισθάς, δίς, δυνατώς, δυσκόλως, δωρεάν, εβραίστι, εμπροσθε, εμπροσθεν, εναντι, ενοδευν, ενθευ, επειτα, εσώ, εσωθεν, ετι, εξω, εξωθεν, ει, ειπερ, ειπως, ελαν, ελευθερ, εφάπαξ, εγγύς, εκ, εξ, εκει, εκείθεν, εκείσε, εκουσίως, ενδοξώς, ενχάδε, ειναιαθα, ειναιευθεν, ειντφος, ειναιφων, επαίρον, επαίνα, επαίνω, επειδή, επει, επιμελος, επι', επ', εφ', ετοιμως, εξαφνης, εξάπτυα, εξης, ευπρεπος, ευβους, ευθεια, ευθεως, ευτόνως, ενεκα, ενεκεν, εως, φανερως, φιλοφρόνως, φρονίμως, γαρ, γε, γουν, ιδεως, ιδικα, ιη, ιητου, ιηθ, ιητη, ινατι, εφως, ιδιφρως, ινα, ιν', καιν, και, καceptors, καθεξης, καθη, καθοτι, καθως, κατα, κατ', καθη, κατεναντι, κατω, κενως, κυκλοθεν, λαμπρως, λαθα, λιαν, μαλλων, μακροθεν, μαλατα, μεγαλως, μενουν, μενουγνυ, μετα, μετ', μεθ, μετασυν, μετριως, μεχρι, μεχρις, μεν, μεντοι, μηδαμως, μηδε', μηδ', μηδεποτε, μη, μην, μηποτε, μητε, μητι, μητη.
The data matrix is therefore of size 99x230.

**Unsupervised Random Forest with PAM Clustering.**

The R packages required were ‘Random Forest’, ‘Cluster’, ‘Mass’. In order to get the best possible results, an unsupervised Random Forest was performed and the proximity matrix of a forest of 800 trees calculated. Using PAM clustering on the proximity matrix, the optimal number of clusters was found. This procedure was repeated 100 times in order to obtain a histogram of ‘best cluster’ values. The forest from each run was combined, giving a total forest of 80,800 trees (the extra 800 resulting from setting up the initial combined forest variable).

The lines of code, with brief explanation, required to do this task are:

```r
> asw <- numeric(25) ; Setting variable for average silhouette width
> best <- numeric(100) ; Setting variable for best number clusters
> for(b in 1:100){set.seed(b*50) ; Setting loop and random number generator
+ data.urf <- randomForest(variables,ntree=800,proximity=TRUE) ; Random Forest
+ for(k in 2:25){asw[k] <- pam(data.urf$proximity,k)$silinfo$avg.width}; PAM loop
+ best(b) <- which.max(asw) ; Selection of best number of clusters
+ data.urfcombi <- combine(data.urf,data.urfcombi) ; Combining forest data
> truehist(best) ; Plot of histogram of best number of clusters
```
The most frequently determined number of clusters is 11.

Therefore, using the combined forest and 11 clusters PAM clustering can be executed with the following instruction:

```
PAMcombi <- pam(data.urfcombi$proximity, 11)
```

The cluster plot is obtained with:

```
clusplot(PAMcombi, main = "PAM Clusplot For 11 Clusters")
```
These two components explain 16.47% of the point variability.

The cluster-plot can best be appreciated when lines are removed and the clusters labelled. Please refer to [Results Table 1] table to identify the cases contained in each cluster.
These two components explain 16.47% of the point variability.

The integrity of the clusters can be found by plotting the silhouette width:

```r
> plot(PAMcombi)
```
Silhouette plot of \( \text{pam}(x = \text{data.urfcombi}\$\text{proximity}, k = 11) \)

11 clusters \( C_j \)

\[
\begin{array}{ll}
1 & 6 | 0.11 \\
2 & 7 | 0.17 \\
3 & 8 | 0.08 \\
4 & 19 | 0.06 \\
5 & 10 | 0.02 \\
6 & 5 | 0.07 \\
7 & 9 | 0.17 \\
8 & 9 | 0.08 \\
9 & 9 | 0.14 \\
10 & 13 | 0.08 \\
11 & 4 | 0.12 \\
\end{array}
\]

Average silhouette width: 0.09

Cluster 5 contains some negative values of silhouette width and suggests that this cluster is of doubtful integrity, that is, it may contain some cases that are mis-clustered. A table of which text samples appear in these cluster is obtained by:

```r
> table(texts, PAMcombiSclustering)
```
Cluster 1 consists of 2 Esdras.

Cluster 2 consists of Gospel of John.

Cluster 3 consists of parts of Genesis and Exodus.

Cluster 4 consists of parts of Genesis and Exodus.


Cluster 6 consists of 4 Maccabees AND Wisdom of Solomon.

Cluster 7 consists of Pauline Corpus.

Cluster 8 consists of Acts and part of Luke (Lukan Corpus?).

Cluster 9 consists of Sirach (Ecclesiastes).

Cluster 10 consists of Mandates, Similitudes and Visions (Shepherd Corpus).

Cluster 11 consists of Revelation.

[Results Table 1]

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>2Es</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4M</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ex</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>6</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>G</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ga</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>IC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>IIC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>In</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Lk</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>M</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ro</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Rv</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>S</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Si</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>So</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>V</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
The method therefore shows that *The Shepherd* corpus is by the same author. We can therefore argue for the single author hypothesis. The method is further substantiated as a discriminatory method because it also identifies the Pauline corpus and perhaps a Lukan corpus. However, Luke and Acts are found split, but together, in two cluster which suggests some editorial work is involved in the compilation of Luke and Acts.

Although the optimum clustering cluster The Pauline corpus as a single cluster and *The Shepherd* corpus as a single cluster, it might be instructive to determine the relationship between the parts of the texts in each cluster. This means looking at higher numbers of clusters. The histogram of best number of clusters clearly shows that 11 clusters is the optimal clustering level, however, for the sake of interest we consider 22 clusters. The robustness of the Pauline corpus is shown by the fact that even this deep into the partitioning it still remains a single cluster.

<table>
<thead>
<tr>
<th></th>
<th>8</th>
<th>19</th>
<th>20</th>
<th>21</th>
</tr>
</thead>
<tbody>
<tr>
<td>2Es</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4M</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ex</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>G</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ga</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>IC</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>IIC</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Jn</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Lk</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>M</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Ro</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ry</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>S</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Si</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>So</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>V</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

*The Shepherd* corpus is split into three. If this structure is NOT random chance at this deep level of partition it suggests that perhaps the compilation of the whole work uses parts of Similitudes as bridging units. Thus Visions links particularly to one case from Similitudes; Mandates links with two cases from Similitudes; four cases
from Similitudes bind the other two sections together. This can only be speculative since we do not know the influence of randomness involved at levels so low in the clustering process.

**Supervised Random Forest**

This method constructs the forest using the data labels i.e. the text label from which the samples are drawn. It then tries to find the combination of variables that will assign the sample to its correct class (text). This is carried out using the Random Forest package in R, with the following command:

```
> data.rf <- randomForest(variables, texts, ntree=5000, importance=TRUE, proximity=TRUE)
```

5000 trees are ‘grown’ and the resulting confusion matrix obtained.

Type of random forest: classification  
Number of trees: 5000  
No. of variables tried at each split: 15

OOB estimate of error rate: 24.24%

Confusion matrix:

<table>
<thead>
<tr>
<th></th>
<th>2Es</th>
<th>4M</th>
<th>A</th>
<th>Ex</th>
<th>G</th>
<th>Ga</th>
<th>IC</th>
<th>Jn</th>
<th>Lk</th>
<th>M</th>
<th>Ro</th>
<th>Rv</th>
<th>S</th>
<th>Si</th>
<th>So</th>
<th>V</th>
<th>class.error</th>
</tr>
</thead>
<tbody>
<tr>
<td>2Es</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4M</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ex</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.4166667</td>
</tr>
<tr>
<td>G</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>14</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.125</td>
</tr>
<tr>
<td>Ga</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>IC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.3333333</td>
</tr>
<tr>
<td>IIC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Jn</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Lk</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.1111111</td>
</tr>
<tr>
<td>M</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Ro</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.6666667</td>
</tr>
<tr>
<td>Rv</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>S</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>Si</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>So</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>V</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

143
The text labels across the top of the table (highlighted in yellow) are, in effect, named clusters. What would be cluster 1 in the previous table is now the 2Es cluster. From the first vertical column (the orange column), the 6 blocks of text from 2Es can be classified into the 2Es cluster. The column of class error indicates how many sample blocks were classified in another cluster. We can ignore the class error for our purposes. Some columns contain nothing but zero values eg. the V column, these are empty clusters. For the sake of ease of interpretation, therefore, the table is reconstructed to show cluster containing samples, and the column labels are changed to the more familiar cluster number. This is purely a presentational change, it does not change the results at all.

### [Results Table 2]

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>2Es</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4M</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ex</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>G</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>14</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ga</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>IC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>IIC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Jn</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Lk</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>M</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ro</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Rv</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>S</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Si</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>So</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>V</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Cluster 1 consists of 2 Esdras samples.

Cluster 2 consists of some of 4 Maccabees and one case of Wisdom of Solomon.

Cluster 3 consists of all of Acts and part of Wisdom of Solomon.

Cluster 4 consists of Genesis and Exodus.

Cluster 5 consists of Genesis and Exodus and one part of Luke.
Cluster 6 consists of 1 part of Romans.
Cluster 7 consists of 1&2 Corinthians and part of Romans.
Cluster 8 consists of Galatians.
Cluster 9 consists of Gospel of John.
Cluster 10 consists of remainder of Gospel of Luke and part of 2 Corinthians.
Cluster 11 consists of 1 part of Similitudes.
Cluster 12 consists of remaining parts of Romans and 1 Corinthians.
Cluster 13 consists of Revelation.
Cluster 14 consists of Mandates, Visions, most of Similitudes.
Cluster 15 consists of Sirach (Ecclesiasticus) and 1 part of Wisdom of Solomon.

Performing a PAM clustering on the proximity data suggests a 10 cluster solution from the average silhouette width plot:
Determination of Number Clusters

A silhouette plot for the 10 cluster solution is:
The silhouette widths are much larger in this analysis than with the unsupervised case, suggesting a better clustering of the data. This is indeed the case, as can be seen from the cluster plot:
Supervised Random Forest with 10 Clusters

These two components explain 23.4 % of the point variability.
Supervised Random Forest with 10 Clusters

These two components explain 23.4% of the point variability.
Cluster 1 consists of 2 Esdras.
Cluster 2 consists of Gospel of John.
Cluster 3 consists of Exodus and 2 samples Genesis.
Cluster 4 consists of 4 Maccabees AND Wisdom of Solomon AND Pauline Corpus.
Cluster 6 consists of Sirach (Ecclesiasticus).
Cluster 7 consists of Acts.
Cluster 8 consists of major part of Genesis.
Cluster 9 consists of Mandates, Similitudes, Visions (The Shepherd corpus)
Cluster 10 consists of Revelation.

Supervised Random Forest with PAM clustering shows that The Shepherd corpus is the work of a single author. However, the method does not separate the Pauline corpus from some other texts. It does not pick out differences in a Lukan corpus. Supervised Random Forest with PAM clustering is not a reliable working
Finite Mixture Density.

The number of variables must be less than the number of cases. Multicollinearity is a concern. In the model building process described in the previous chapters, the variables were selected from the main set by choosing those that had the largest means. It took a great deal of time and manipulation to find the best variables and avoiding multicollinearity. A better method involves the use of random forests with a selection technique to drop the least important variables on each run. The software to do this is varSelRF for R. The following lines of program will perform the operation on the data already collected. The method works by supervised random forest.

```r
> rf.varsel<- varSelRF(variables, texts, ntree=5000, ntreeIterat=2000, vars.drop.frac=0.2, whole.range=TRUE)
```

The important variables chosen\(^{107}\) by this method were:

\[\begin{align*}
\text{ίνα, ίν', καί, μετά, μεθ', μεθ' μέν; μή, ού, ούχ, ούκ, οῦν; οῦ; άλλα, άλλ', πρό, προ', προθεν; σύν, ύπό, ύπ', ύφ', ώ';}
\text{ἀπό, ἀπ', ἀφ', δὲ; διά, δι', εἰναντί; ἐκ, ἐξ; ἐκεί, ἐκείθεν;}
\text{ἐπί, ἐπ', ἐφ', ἐώ'; γάρ; ἡμίκα;}
\end{align*}\]

A reduced data matrix consisting of only these variables was transformed by the function SQRT(data)+SQRT(data+1) and EM clustering performed.

```r
> data.em<-EMclust(SQRTdata,G=1:25)
> plot(data.em)
```

The resulting plot is shown below:

\(^{107}\) These are not the same as the variables chosen if the 24 highest frequency words were used. In fact, if the top 24 highest frequency words were used instead, the clustering partitions are found to be worse and only 9 clusters found. The varSelRF variables finds 10 clusters, although the integrity of some clusters is doubtful.
The best model is C or D with 9 to 12 clusters. The differences in BIC between these models is negligible. A 10 cluster "VEI" model was estimated:

```r
> data.em3 <- summary(data.em, SQRTdata, G=10, modelNames="VEI")
> table10 <- table(texts, data.em3$classification)
```

The following table of classification was obtained:
<table>
<thead>
<tr>
<th>texts</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>2Es</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4M</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ex</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>7</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>G</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ga</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>IC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>IIC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Jn</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Lk</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>M</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ro</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Rv</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>S</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Si</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>So</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>V</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Cluster 1 consists of 2 Esdras.
Cluster 2 consists of Gospel of John.
Cluster 3 consists of Exodus and Genesis.
Cluster 5 consists of most of Luke with one case from Exodus.
Cluster 6 consists of 4 Maccabees, **Mandates, Similitudes, Visions** and Wisdom of Solomon.
Cluster 7 consists of Galatians, 1&2 Corinthians and Romans (Pauline Corpus) and one case from Wisdom of Solomon.
Cluster 8 consists of Acts and one case from Luke.
Cluster 9 consists of Sirach (Ecclesiasticus)
Cluster 10 consists of Revelation.
Discussion

The results from all four tables show that the Mandates, Similitudes and Visions always cluster together. The only exception is Results Table 2 which separates one part of the Similitudes from the remainder and places it in a cluster on its own. Table 2 is the only one that suggests a large number of clusters for the data. PAM clustering indicates that the ‘true’ number of clusters is more likely to be 10 or 11, hence the supervised confusion matrix is an over-clustered solution. This over-clustering also breaks-up the Pauline corpus in a similar manner, with some clusters only containing one sample.

Whether or not over-clustering has produced questionable results, it can be seen that the majority of The Shepherd corpus is clustered together. The inevitable conclusion of this data is that The Shepherd is probably the product of a single author.

The word ‘probably’ is included because statistical analysis never gives absolute results. Statistics is based on probability, therefore the result it produces are probable results. It is possible to say with a high degree of certainty, however, that the single authorship solution is correct and more probable than any of the multi-authorship theories.

The most difficult explanations are related to the interactions of Genesis, Exodus, Luke and Acts. Genesis and Exodus are clustered over a number of groups, with some parts of both texts in the same groups. This is not surprising and indicates that the author’s hand is not the only one, but the text has been thoroughly revised by editors. Acts and Luke are the surprising texts. Acts mostly groups its 8 samples in one cluster. Parts of Luke are found in the Acts cluster as follows:

Results Table 1 Luke and Acts are split into two clusters with one of those clusters containing the majority of Luke, one case from Acts and one case from Exodus. The other cluster (which may be the Lukan corpus) consists of the majority of Acts and two cases from Luke. These are interesting connections and deserve more attention than can be given in this thesis.

108 The connections would need to be considered in the light of the different genres found in these two texts, speeches, narrative, travel diary etc. Also the question of Luke’s use of ‘Septuagint style’ in his work.
Results Table 2 has Luke and Acts in separate clusters. One case of Luke appears clustered with substantial parts of Genesis and Exodus.

Results Table 3 has Luke and Acts in individual clusters of their own with no other cases present.

Results Table 4 has 1 sections of Luke and 8 sections of Acts in the same cluster (a possible Lukan corpus). The majority of Luke clusters with one case from Exodus. Finally, one case of Luke clusters with the majority of Exodus and Genesis.

While there is a possible Lukan corpus consisting of Luke-Acts, the proportion of Luke involved is not certain, it could be as little as 2000, or as much as 4000 words. What is causing the variability of Luke in the clustering process? It appears to be related to parts of Genesis and Exodus. However, where Luke appears in clusters with Genesis or Exodus, it is usually with small numbers of cases of the other, for example 7 cases of Luke with only one case from Exodus, or, the majority of Genesis and Exodus with only one case of Luke. Somehow, the clustering algorithm is sensitive to Luke-Genesis-Exodus. Because sections of Luke are so similar to sections of Exodus and Genesis, slight changes in the clustering criterion for one has dramatic affects on the other two. Mealand also had difficulties with Luke. His explanation was that Luke wrote in ‘Septuagint style’. This invalidates Mealand’s methodology because it suggests the method measures style rather than authorship. The same argument would therefore apply to the work of this thesis. Is there another explanation? The sections of Luke that distort the Luke-Acts corpus theory is the section containing the ‘birth narrative’. This section may act as a ‘chaining link’, a type of problem well-known in Cluster Analysis, where a single (sometimes more) data point lies between two close clusters and acts to combine the clusters sooner than would otherwise be expected:
In the illustration on the left, two clusters are well separated. In the illustration on the right, the one chaining sample combines with a cluster to form a new cluster (red ellipse). This new cluster is now closer to the other cluster than would be expected, simply because of the one data point. Such chaining would explain why the problems arise with Luke and parts of Genesis and Exodus\textsuperscript{109}. If one of the clusters in the diagram above was sections of Luke and the other sections from Genesis and Exodus and the chaining point was the birth narrative data, then we can understand how parts of Luke get ‘wrongly’ clustered with Genesis/Exodus. This raises the obvious question, “where did Luke get his source of the birth narrative?” This is not a question for this thesis to pursue, it is only raised as a possible explanation of the Luke-Genesis-Exodus problem in defence of the methodology used in this chapter.

This chapter has shown that a statistical method for authorship detection has succeeded in producing good and interpretable results using Random Forest and Pam clustering, backed up with Finite Mixture Density analysis. It has suggested a single authorship solution for *The Shepherd*. Using that knowledge we can proceed to part two of the thesis and a continuing grammatical analysis of *The Shepherd*.

**Conclusion**

The aim of the first part of this thesis was to try to resolve the problem of the number of authors involved in the production of *The Shepherd*. Opinion has always been divided between the two basic options, either a single author named Hermas, or

three different authors, as proposed by Giet. There was a third option, the multi-author option of Coleborne, but no one, other than Coleborne, thought that a serious option.

Building on Mealand's successful statistical analysis of Luke-Acts, it was decided that the best way to solve the authorship dilemma was by the use of a statistical analysis of the function words found in The Shepherd. Such an analysis had not been attempted, other than Coleborne's effort in the 1960's. In the previous chapters I have shown that the most widely used statistical procedure of Z-transformation, PCA and cluster analysis using Wards' method, are not the most reliable methods to use.

PCA suffers from the fact that if different variables are used, different results will be obtained. Variable selection is a therefore a task of great importance, and great practical difficulty for the text analyst. One badly chosen variable can affect the results produced and the conclusions drawn.

The variables measured as word counts will also be taken from mainly skewed data. PCA does not perform well with skewed data, therefore the Z-transformation (it is usual to perform PCA on the correlation matrix) is always used to try and resolve the problem. However, rather than resolve the problem, it can make matters worse.

Random Forest algorithms are the way forward. The method does not require transformed variables, for it will work equally well with skewed and normal data. Even though forests of several thousand trees may be 'grown', the method does not over-fit the data.

The statistical analysis concludes that The Shepherd was written by a single author. We are not able to conclude from the statistical analysis alone, whether it was written over an extended period of time, nor can the suggestion of a date be proposed for its time of writing. These questions need a different type of analysis than the one performed in this statistical section and will be answered in Part Two of this thesis, where a different methodology is developed specifically for that purpose.

In the introduction a number of questions were posed for this thesis to answer. Not only was the number of authors to be determined, but the following required comment:

*Do Visions 1 to 4 show evidence of a different author to that of the Mandates and*
Similitudes?

This question was asked because of the evidence that the Visions were circulating in the early church as an entity in their own right, separate from the Mandates and Similitudes. The results of the statistical analysis show that the answer to this question is that there is no evidence that Visions 1-4 were written by a different author. Statistics cannot comment on whether the Visions were written first and put into circulation before the remainder of The Shepherd was written, all that the results will allow us to say is that the author of Visions is the same as the author of Mandates and Similitudes.

Is M12.3.3 to M12.6.5 written by a different author/editor?

It has been shown that the methodology outlined and used above is sensitive enough to be able to detect editorial that may be included in the text of The Shepherd. There is no evidence that this editorial seam was written by another author. Statistics cannot answer the question whether or not this section is an editorial seam written to join the Mandates and Similitudes into a single literary unit, but the results suggest that if this was the case, it is the same author who has written both sections who also writes the seem.

Do the Similitudes display the same characteristics of authorship as the Mandates?

The results of the statistical analysis suggest that Similitudes and Mandates had a common author. Statistics cannot answer the question whether Similitudes and Mandates were written at different periods in the author's career, only that they display the same characteristics of authorship.

Does Similitude 5 show evidence of a different author or the presence of editorial links?

To repeat all that has been noted so far, Similitude 5 does not display any different authorial characteristics to the rest of The Shepherd. Any structural relationship that may exist between this Similitude and the rest of the text is that of the single author's craft.

Does Similitude 9 show evidence of a different author or the presence of editorial links?

The same comments apply here as in the previous question. Single authorship
is the key.

The statistical analysis, developed in this thesis, has produced confirmatory evidence for the single author hypothesis. It also suggests that the author, Hermas, was not 'as ordinary' as some commentators would suggest. His use of language and grammar has been suggested to be that of the average person. However, he appears to have constructed The Shepherd, including editorial links, entirely on his own. This would suggest that he was an expert in his craft of writing.

All that this thesis set out to prove regarding authorship has been successfully achieved by the use of Random Forest and PAM statistical methods and using a FMD technique as confirmatory back-up. Not only does the statistical methodology work, it works well and provides answers to problems that no one else has been able to answer.

The next part of the thesis continues the analysis of the grammar of The Shepherd in order to discover the problems and difficulties that Hermas and his community faced in Rome in the early 2nd century.
Chapter 10

MAJOR THEMES IN THE SHEPHERD:
SIMPLE LEXICAL PATTERN

Introduction

As mentioned in the last chapter we now enter the second part of this thesis which deals with the social location of Hermas and his community. A different methodology to the statistical one of the previous chapters will need to be developed to meet this new challenge. This chapter explains the methodology that may be used in order to uncover the information encoded in a text by the writer. We will begin by looking at Vision 1 and develop a method that allows us to extract information about the themes it contains and its literary structure. There will be a number of constraints upon the method developed to do this task, in particular, the method must be general in its scope so that it may be applicable to any text. The rewards of analysing the construction of the discourse, and the stance the author takes on major issues, are that we will be able to understand the concerns and problems of the sub-group of society to which the writer belongs. By noting the grammatical methods used by the writer, we will be able to assess his power of persuasion, and, the focus and direction of his leadership.

Discourse Construction

A text is more than a collection of words arranged by some underlying set of grammatical rules. It is an expression conveyed by meaningful glyphs, of the thoughts, feelings, aspirations, and, perhaps, the social location of its author. By identifying what the writer thinks, we are able to reveal the ideology of the society in which the writer works. This ideology reflects, of course, not the whole society, but rather the sub-group to which the writer belongs. Therefore, the text expresses the opinions of the writer about the events that are reported. These opinions may be openly stated in the form 'here is what happened, and this is my opinion concerning the matter'.

However, the opinion of the writer may be expressed in subtler ways signalled by the use of discourse markers. Various discourse markers could be used, the choice being dependent on the rhetorical intention of the writer. For example, the writer may express certainty - 'here is what happened, obviously it means...'. On the other
hand, perhaps the writer may wish to express a distance or detachment from the statement - 'here is what happened, it is claimed that it means..'. The word 'obviously', and the phrase 'it is claimed that', are the discourse markers that reveal the writer's attitude to the initial proposition or statement. However, they also bring the reader into a relationship with the writer by the sharing of a common ground. Not only does the reader share the common ground of meaning - 'obviously it means', the reader also supplies the information that substantiates that common ground by making a positive mental evaluation of the claim. The writer is therefore able to build, and maintain, a relationship with the reader in an attempt to persuade them to interpret the facts in a particular way.

The first part of this task is to consider Vision 1 as a narrative told by the writer, who by implication, is identified as the main character of the story, Hermas.

Labov identified a six-part structure to a fully formed narrative:

The Abstract. This summarizes the main point of the story in advance.

The Orientation. This section provides information on the setting, giving details of place and time.

The Complication. This is the main part of the narrative that describes the problem, or sequence of events that gives the interest to the story.

The Resolution. A description of the outcome of the Complication.

The Coda. The conclusion of the narrative.

The Evaluation. The attitude of the writer to, and explanation of, the events described in the narrative.

Not all of these parts are necessarily found in a narrative. A writer can, however, increase the complexity of the narrative by recursive use of sections 2, 3 and 4.

Vision 1 begins with an abstract. V1.1.1&2 summarizes the point of the story, namely the thought - "I should be happy if I had a wife of such beauty and character". It is this seemingly innocuous thought that is the discussion point for all that follows in the narrative.

The orientation is provided in verse V1.1.3, "After some time, while I was

going to Cumae". Although the original text is unclear, a choice between ἐλθὲ Κώμας (to the countryside), and ἐλθέ Κούμας (to Cumae, a town 130 miles distant from Rome), the purpose of the verse is to set the narrative in a spatial setting.

The complication consists of the vision of Rhoda, who, by this point in the narrative, appears to have achieved heavenly status. In this part of the narrative Hermas is told the true nature of the problem - "God is angry with you because you sinned against me" (V.1.6).

The resolution describes the abrupt end to the vision which leaves Hermas in a fearful and depressed state - "After she had spoken these words the Heavens were shut, and I was all shuddering and in grief" (V.1.2.1).

In the coda, or conclusion to the first part of the narrative, the writer leaves Hermas pondering - "with what words shall I beseech the Lord to be forgiving unto me?" (V.1.2.2).

The narrative now takes a recursion, defining a new orientation - "While I was considering and doubting these things in my heart" (V.1.2.2). A new character is introduced, a vision of an old woman, along with a new complication - "But it is not for this that God is angry with you, but in order that you should convert your family" (V.1.3.1). The new resolution leaves Hermas in a better state of mind - "Lady, this last part pleases me, but the first part was hard and difficult" (V.1.4.2), and the vision does not end abruptly, the woman is still speaking as she disappears from sight. The coda is a call to courage - "Play the man Hermas" (V.1.4.3).

From the above, we can see that Vision 1 displays a typical narrative structure. Two sub-narratives are fused together in order to make the whole, however, these two narratives are not merely two adjacent stories. Ryave suggests that in multiple-story narratives, the second story endorses the first and builds on it. In fact, Ryave suggests that each succeeding story exaggerates the one before it and the later stories become an evaluation of the earlier stories. Admittedly, Ryave's work was related to oral story telling rather than a written narrative, but there is evidence of exaggeration in the narrative of Vision 1 (and perhaps evidence for an original oral narrative). In the first story God is angry with Hermas because of his personal adulterous thought,

in the second story the focus widens to the sins of his family.

From the point of view of discourse analysis, the claim of the Vision Character, Rhoda, is a hidden averred statement, that is, the statement, "God is angry with you", is evaluated by Hermas (and the reader) as, "God says He is angry with you". Implicit in the statement is the assumption that God is the source of the statement (attribution). However, the source of the statement is not God but Rhoda (averral). The writer forces his readers to accept this opinion by using a grammatical 'sleight-of-hand'. He achieves this by:

Qualifying the subject of the predicate with a series of three participial phrases, "... who dwells in Heaven I and created that which is out of that which is not I and increased and multiplied it for the sake of His holy church... ". This places stress on the subject, heightening the importance of the subject in the mind of the reader, separates the subject from the predicate and thus obscuring the true source of the statement in the narrative.

The author exaggerates Hermas' evaluation of the statement, giving him physical, "I was all shuddering and in grief (cf. Daniel 7:15), and, mental anguish "How shall I be saved?", "How shall I propitiate God?", "How shall I beseech the Lord to be forgiving?". The reader is encouraged to accept the validity of the statement because of the effect it is seen to have on Hermas. Following the heightened superiority of the subject, we are given the heightened inferiority of Hermas. The reader observes a reassuringly human Hermas; one who reacts to the accusation in a manner they would probably adopt.

By a skilful use of a grammatical conjuring trick and a play on emotions, the writer encourages the reader to believe that the statement, "God is angry"... is true. This acceptance is vital to the writer because all that follows in the Visions, Mandates and Similitudes is based on the truth of this statement. In addition, the acceptance of the validity of the statement implies the validity, and authority, of the character making the statement. The writer needs to convince the reader that the Visionary Character (hence the visionary) has a high status as 'information provider'. The writer is so successful in achieving the desired result that the reader accepts a change of emphasis of the second Visionary Character over the first without questioning the authority of the first.

A relationship between writer and reader is formed through the weak and human character of Hermas who is seeking enlightenment, and the authoritative
Vision Character who offers that enlightenment. This can be shown diagrammatically:

\[
\text{READER } \leftrightarrow \text{ HERMAS } \leftrightarrow \text{ VISION CHARACTER } \leftrightarrow \text{ WRITER}
\]

The quest for understanding is played out in the interaction between Hermas and Vision Character and it proceeds via a question and answer format (V1.1.7&8). The question and answer format is basic to our understanding of the whole work of The Shepherd. The writer informs and influences his reader via the dialogue between the narrative characters.

The writer stands in similar relation to the reader as the Vision Character to Hermas in the narrative. The Vision Character speaks on behalf of God; therefore, the writer also assumes authoritative status.

Having established the authority of the Vision Character, the discourse about 'the righteous man' (V1.1.8-9) then assumes a like authority and significance. A number of moral and ethical issues are evaluated in this discourse:

<table>
<thead>
<tr>
<th>What is Evaluated</th>
<th>How it is Evaluated</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 an evil desire</td>
<td>it is an evil deed</td>
</tr>
<tr>
<td>2 an evil deed</td>
<td>it is a sin</td>
</tr>
<tr>
<td>3 righteous man</td>
<td>has righteous designs</td>
</tr>
<tr>
<td>4 righteous designs</td>
<td>his repute stands fast</td>
</tr>
<tr>
<td>5 righteous man</td>
<td>assisted by the Lord</td>
</tr>
<tr>
<td>6 evil designs</td>
<td>bring captivity and death</td>
</tr>
<tr>
<td>7 unrighteous man</td>
<td>no hope and abandoned</td>
</tr>
<tr>
<td>8 you (Hermas) pray</td>
<td>shall be healed, and house and saints</td>
</tr>
</tbody>
</table>

The desire and the deed are alike regarded as sin. The theme of the discourse is that the righteous man will have righteous designs (intentions) resulting in positive outcomes. Those who have the wrong attitude (evil designs/intentions) are displayed negatively using the strong language usually applied to prisoners of war - captured, hopeless, abandoned, death.

The discourse of Rhoda concludes with a call to Hermas to pray (προσεύχου - imperative mood).

However, the purpose of the first part of the story is to lay the foundation for the more important Visionary Character, the 'old woman' (later identified as a personification of 'the church'). The reader is now prepared to accept the authority of
this new character, and the story is broadened to encompass a wider theme, that is, Hermas' responsibility for correcting the sin of his family. The 'old woman' gives a longer discourse than Rhoda thus emphasizing the importance ascribed to this second Visionary Character.

Following the preamble, the discourse of the 'old woman' shows three interesting features.

1) **A progressive argument making use of the adversative conjunction.**

Having established that Hermas is gloomy because of his sinful thoughts, the discourse that follows (V1.3.1-4) is remarkable for its high frequency of the word ἀλλὰ, 'but'. The average frequency of this word across a wide range of texts is 5.5/1000 words\(^{112}\); the average frequency in *The Shepherd* is 4.9/1000 words. Compare these values with the 7 times the word is used in only 99 words of discourse in V1.3.1&2 (equivalent to 70/1000 words!).

<table>
<thead>
<tr>
<th>Use of 'but' in V1.3.1&amp;2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 But it is not for this that God is angry</td>
</tr>
<tr>
<td>2 But in order that you should convert your family</td>
</tr>
<tr>
<td>3 But you are indulgent</td>
</tr>
<tr>
<td>4 But have allowed them to become corrupt</td>
</tr>
<tr>
<td>5 But He will heal</td>
</tr>
<tr>
<td>6 But the great mercy of the Lord has had pity on you</td>
</tr>
<tr>
<td>7 But have courage</td>
</tr>
</tbody>
</table>

A clause beginning with 'but' can perform a number of functions. It can tell the reader that what follows is not what is expected\(^{113}\). This is certainly true of the first clause in the table above. Rhoda had stated that God was angry with Hermas because of his adulterous thought, the 'old woman' now seems to contradict that claim, "But it is not for this". This prepares the way for new information to be introduced into the story. She does, however, affirm the initial proposition, "God is angry with you".

---

\(^{112}\) Obtained from the statistics provided by Perseus. The following statistics might be of interest, all are reported as frequency/1000 words:
*Acts = 1.6; Luke = 1.8; Hebrews=3.2; Romans = 9.7; Galatians=8.1; Ephesians = 5.4; Mark =4.0; Genesis = 0.8; Exodus = 0.4; Jeremiah = 0.8*

The second 'but' clause modifies the main clause and introduces a purpose clause. The new information is now introduced to the reader and the focus turns to Hermas' family, which has sinned against the Lord.

In the second sentence of V1.3.1, the accusation is laid against Hermas, "You do not correct your family. The sentence is introduced with the 'but' clause (number 3 in the table above) modifying the subject, Hermas, in a negative sense as being a φιλότεκνος that is, an over indulgent father.

The 4th 'but' clause simply contrasts clauses in the form, "You do not correct... but you have allowed...". The contrast is between what the conduct should have been and what it actually was!

The third sentence in V1.3.1 repeats the claim of anger, but with slightly different words, "The Lord is angry with you".

The 5th 'but' clause simply contrasts clauses, contrasting a current state with a future state.

The 6th 'but' clause opens V1.3.2 and clearly contrasts the current state of Hermas' family (sin, wickedness and corruption) with a future state (strong, established).

The 7th 'but' clause has the grammatical form of ἀλλά followed by a verb in the imperative. The 'but' strengthens the command "have courage".

Drawing a sentence diagram\(^{114}\) of these two verses clarifies the above statements and introduces us to the writer's method of producing coherent narrative structure.

\(^{114}\) Rules for sentence diagramming may be found in: Stevens,G.L. New Testament Greek. Procedures for sentence diagramming may be found in: Fee, G.D. New Testament Exegesis.
V1.3.1

'Άλλη·

οὐχ ἔνεκα τούτου

ὁ θεός ὁργίζεται σοι

ἀλλ'·

ἵνα (γι' σου) ἐπιστρέψῃς τὰν οικόν

σου

τὸν ἄνθρωπον

eἰς τῶν κήρυκων

καὶ

eἰς ἰμάς

tοὺς γονεῖς

αὐτῶν

167
V1.3.2

άλλα

ἡ πολυπλαγχνία ἤλεησέν σε

τοῦ κυρίου καὶ

καὶ τὸν ὄικον σου

ἰσχυροποιήσει

καὶ 

θεμελιώσει 

ἐν τῇ δόξῃ αὐτοῦ

μόνον

μὴ ἰδαμήσεις

ἀλλὰ 

εὐφύξει 

καὶ τὸν ὄικον σου

ἰσχυροποιεῖ
ἀλλά

ὡς φιλότεκνος

(you) ἐνοθέτεις τῶν οἴκων

οἶκος σου

ἀλλά

ἀφίξες αὐτῶν καταφθαρήματι

διὰ τοῦτο

ὁ κύριος ἀργίζεται σου

ἀλλά

ἰάσεται σου

πάντα τὰ προγεγονότα ποιημὰ

γὰρ ἐν

σὺ κατεφθάρῃς τῆς οἴκης σου

διὰ ἀπὸ τῶν βιωτικῶν πράξεων

τὰς ἁμαρτίας

καὶ ἔκείσσων

ἀνυμήματα
VI.3.2

Not only can we see the function of the ἄλλα phrases, but we now see a number of interesting structural signals become apparent (note the highlighted words). We will come back to these later in this chapter.
2) The use of a supporting image.

The rather elaborate argument of the discourse is interrupted with an illustration taken from common experience, that of a smith working metal.

This change from argument to personal, situational experience not only helps to draw this section of the discourse to a conclusion, it also acts as a bridge into the next section. Hermas is encouraged to patiently continue the work of correcting his children. This section closes with a conditional clause, "if they repent with all their heart, they will be inscribed in the books of life with the saints."

Mention of "the books" takes us back to the beginning of the vision, "and there came a woman...with a book in her hand" (VI.2.2), and this, in turn prepares the reader for the final part of the discourse, "Would you like to hear me read aloud?" (VI.3.3).

3) The use of liturgical material

Osiek\textsuperscript{115} suggests a possible liturgical context for this section (VI.3.4). She notes a common Wisdom theme, some Jewish liturgical expressions and possible Stoic influence. This section of the discourse is certainly different from all that has preceded it. Its theme of praise to God contrasts sharply with the condemnation of Hermas for his failure. The literary form of the section seems to be that of a confession of faith or a catechetical statement of belief. However, this section also performs an important role in the narrative of Vision 1, reminding the reader of the beginning of the visionary experience and uniting various themes by common lexis.

Method Development

It would be useful to have a method that could be used on the text of Vision 1 in order to produce the results given above, and be general enough to be used on any text. We have noted above that when a sentence diagram is drawn it is possible to see more clearly the way in which the writer constructs the narrative. The writer links sentences using recurring phrases, or sometimes, single words. Repetition of words or phrases gives coherence and structure to the narrative, allowing the reader to identify a 'flow' of narrative from one paragraph to the next. A disjointed narrative

\textsuperscript{115} Osiek, C. \textit{The Shepherd of Hermas}. p.50.
that gives the reader few clues to understanding the point of the story can be
disconcerting and tedious.

Repetition of words and phrases is also useful in order to reinforce the point
the writer wishes to make. Many of these repetitions are natural unforced responses
in the narrative making process. They occur because the writer develops various
themes during the narration of the story. As we have seen in the sentence diagram
above, these repeated words tend to be nouns, verbs and participles. In order to
detect theme development and narrative flow, we need to use words that are open
class words.

Michael Hoey\(^{116}\) has suggested a method of analysing the text for repetition
which uncovers the themes running through that text. Themes can begin anywhere
in the text and may run for long distances through it. They may be written as
intermittent ideas that surface only occasionally through the text, but they can be
traced. Theme words, by their continuous occurrence, form 'links' between
sentences\(^{117}\). This can be illustrated using Vision 1 and the theme of 'creation'/create'.

Hermas ponders the creation of God (VI.1.3), the Visionary Character,
Rhoda, describes God as the one who created (VI.1.6), and the Visionary Character,
the old woman, reads a description of God the creator (VI.3.4). Thus, a theme, God
as creator, runs through the main sections of the text. However, is it an important
theme in the narrative? Hoey suggested that any sentence was considered an
important theme indicator if it had enough links with other sentences in the narrative.
Hoey found that he needed 3 to 5 links before he regarded sentences to be bonded.
Sentences that are bonded are good indicators of themes. In a complex narrative,
themes can be inter-twined, or, can merge to produce a new theme.

Hoey was dealing with English language texts and found it necessary to
define complex rules as to what constituted a link. These rules included the syntax of
the words being considered for linkage. According to such rules, the noun 'creation',
used in the example above, would not be linked to the verb 'create'. This causes few
problems when dealing with a Greek text since the two forms have different stems,
however, Greek nouns and adjectives are inflected words and there is no reason to
consider that the different inflections mean different entities, since the inflection

\(^{116}\) Hoey, M. *Patterns of Lexis in Text.*
\(^{117}\) Guthrie (1991) used a similar method in the analysis of The Epistle to the Hebrews. He used the phrase
'hook words' which bind together units of discourse. For details of the approach, see: Lane W. *Hebrews 1-8.*
pp.xc-xcviii.
depends on the grammatical role of the word in the sentence. Therefore, a modification of Hoey's method could be based on the Greek word stem.

Adjectives are treated simply by a modified Hoey method. Although a 'beautiful woman' is a different entity to a 'beautiful creation', a writer who is emphasising the theme of beauty can qualify anything as 'beautiful'. Adjectives can be linked regardless of accompanying noun or noun phrase.

Difficulties arise when linking verbs, however, since the personal suffix will need to be taken into account. Thus the word λέγω can only be linked to another λέγω if it refers to the same subject. Therefore, the Greek text of The Shepherd needs to be grammatically tagged and the additional information concerning the word in its specific context needs to be assigned.

Hoey linked personal pronouns according to the person they represented. In a narrative that has many characters, this distinction may be important. In Vision 1 most of the narrative is dialogue and inclusion of personal pronouns will severely distort the theme. Consequently, in this study, personal pronouns were ignored as indicators of theme, in the development of a working method.

Using this method, linking between nouns, adjectives, verbs and participles were considered. While Hoey used the sentence as a unit of reference, this study will use the verse as unit of reference. This is for the sake of convenience and ease of 'mapping' the themes through the text.

When analysing the results, if a verse contains three sentences, the link may be with only one sentence from that reference. Care must therefore be exercised when explaining the theme structure. The number of links per bond was chosen at a level high enough to show structure, but not so low as to complicate the 'map' and make it difficult to interpret. In the specific case of Vision 1 the bonds / link was set to 4.

A number of data manipulations were required to obtain the matrix shown in appendix. The steps involved were:

- Query the text database for words of interest.
- Crosstabulate to obtain word counts per reference. Delete word counts = 1 (nothing to link to!).
- Matrix multiply columns to obtain links between verses.
- Draw 'map' of structure showing a specified number of links.

When this was done for Vision 1, the following 'map' was obtained.
The width and colour of the arrows is proportional to the bond strength.

The diagram 'maps' the major themes of the writer. A number of points need to be clarified.

- Reference boxes shown in red outline have arrows pointing to them, and none pointing away. These references are theme terminators.

- Reference boxes shown in green outline have arrows pointing away from them, and none to them. These are theme initiators. They begin the lexis that is to be followed through the text.

- Reference boxes that have double-line borders are those verses that have strong word links within the verse.

- Reference boxes shown in gold outline have arrows pointing to and from them. These references act either as bridges that continue the theme through the text, or, as theme transformers. A theme transformer is the closing of one theme and the opening of a new theme. The difference between a bridge and a transformer reference needs to be decided from the context and the link words forming the bond.

- Note also that the map displays a fragment consisting of a bond between two sections only. The major theme of this fragment is independent of any others.

If a bond of lower strength is used, it is possible to find sub-themes that act as weak links uniting these major themes. This forms a more complex map and the major themes get lost among the sub-themes. At this stage of the analysis, we are concerned only with major themes but that does not imply that sub-themes are any less important.

Looking at the fragment link V1.1.4 and V1.2.2, it can be noted that the theme is that of, "Appearance of Visionary Character". The two incidents are described in a similar fashion. Hermas sees (βλέπω) the female Visionary Character who greets him (ἀπας ομαί) using the same greetings formula (Ερμᾶ, χαîρε).

The larger net of bonds is more complicated, but can be treated piecemeal. V1.1.8 seems to be an important theme initiator with links to several other parts of the text.

118 βλέπω is used more than 70 times in The Shepherd and refers to Hermas' 'seeing' in his vision. However, since most of the narrative involves visionary experience, it is difficult to attach any significance to this fact.
vision, two of these bonds are the strongest ones found in the whole analysis. The
two strongest bonds are between VI.1.8 and VI.2.4, and between VI.1.8 and VI.3.2.
VI.1.8 also has a number of internal links. Because of its importance as a theme
initiator, this verse acts as a good summary of the whole narrative. VI.3.4 has five
bonds terminating and has internal links. Because of its importance as a theme
terminator, this verse acts as a conclusion to the narrative.

We can discover the major themes developed by the writer by noting the words
that form links between the sections of interest. It is also possible to note how these
themes are structured to form the narrative. If we begin with the lower part of the
map, we can see that the writer compares two contrasting conditions.

```
V1.3.1  Abandonment/Corruption
         ↑
  V1.1.9
       ↓
V1.3.2  Repentance/Life
```

These contrasting outcomes form a conclusion to themes in VI.1.8 and VI.2.4.
The bottom part of the 'map' displays the structure of theme development in the
narrative.
V1.2.4 is a restatement by the second Visionary Character of the discourse uttered by the first Visionary Character. There is a slight difference in emphasis between the two statements. In V1.1.8, the wicked desire (ἡ πονηρά ἐπιθυμία) is regarded as an evil deed (πονηρόν πράγμα), and both the desire and the deed are regarded as sin. The righteous man, however, finds acceptance ("he finds- the Lord ready to assist him"), but the unrighteous man finds death. V1.2.4 changes the πονηρόν πράγμα of V1.1.8 to a πονηρόν ἐργον preparing the reader for the picture of the smith about his work.

The theme seems to be a practical illustration of Matthew 5:27,28.

Πᾶς ὁ βλέπω γυναῖκα πρὸς τὸ ἐπιθυμῆσαι αὐτὴν ἢδη ἐμοίχευσεν αὐτὴν ἐν τῇ καρδίᾳ αὐτοῦ.

Hermas' ἐπιθυμία is described as an evil πράγμα (V1.1.8) and as an evil ἐργον (V1.2.4), so that thought and deed are alike. The realization that 'God sees the heart', and judges thought equally with deed, terrifies Hermas,

Κάγω ὅλος μηνπεφρικός καὶ λυποῦμενος (V1.2.1).

Without repentance, there is only abandonment. Evil desire that rises up in the heart needs to be overcome by repentance with all the heart. These five verses form a theme that can be broadly described as, 'Evil desires and how to overcome them'.
The theme 'map' also shows V1.1.6 as a theme initiator. Looking at each bond in turn, beginning with:

V1.1.6 and V1.1.7 suggest a courtroom scene and deals with the accusation and the cross-examination. A concerned Hermas asks Rhoda (the Vision Character), "Are you now accusing me?" ie. setting out the words (implied) of accusation. "No," she said, "but listen to the words which I am going to say to you. 'God... is angry with you because you sinned against me'." Hermas then begins the cross-examination, "Did I sin against you? In what place, or when did I speak an evil word to you? ... Why do you charge me falsely...?"

Next, the bonded references:

V1.1.6 and V1.2.1 display Hermas' isolation. V1.2.1 begins, "After she had spoken these words the heavens were shut...". This does not simply mark the closing of the vision. The Visionary Character has reminded Hermas (V1.1.6), "God ... dwells in heaven ... ." The closed heaven symbolizes Hermas' isolation from God. He then contemplates what words of request will bring forgiveness, "Or with what words shall I beseech the Lord to be forgiving to me?".

Next, the bonded references:
V1.1.6 and V1.3.3 involve a change of Visionary Character but the idea of listening to the words uttered by the character is present. The first character commands Hermas to listen to the words she is going to speak, the second character commands Hermas to listen to the words she is going to read. Hermas describes his response to the words of the reading in V1.3.3 and that theme is continued into V1.4.2 when Hermas is asked, "Did my reading please you?" V1.3.3 acts as a bridge, continuing the theme into the next reference.

This theme, centred on the command to Hermas to listen, is a theme dealing with authority structures. Both of the Vision Characters inform Hermas of his standing in the sight of God and the reasons why God-Hermas relations are the way they have described. The lines of demarcation of authority are clear, Hermas is subordinate to the Visionary Characters. As 'spiritual beings'\(^{119}\) they have knowledge that ordinary mortals like Hermas (and his readers) do not have. The reading of the book is another display of authority. The one who has the book and reads the words commands the attention of the audience. This very important theme will be considered again later in this chapter.

The reference map also displays the importance of V1.3.4 as a theme terminator by the number of references that point to it.

\(^{119}\) In the sense that they are not physical beings but psychological entities, whether real or imagined.
The callout boxes summarise the theme of the reference bonds. V1.1.6 forms the strongest bond with V1.3.4, the latter being a more elaborate statement of the former. The terms 'holy church' and 'creation' are the major terms in this bond. The church is God's creation and is specially blessed. Similarly, V1.1.8 singles out the righteous/chosen as the recipients of the assistance and blessing of God.

V1.2.4 is a tentative link that contrasts evil intention, which brings sin on the servants of God, with the intention (glorious counsel) of God to create what is beautiful and beneficial for His chosen ones.

V1.3.3 bonds with V1.3.4 since the latter explains the reading of the words that were 'profitable' and 'gentle'.

V1.1.3 links the vision scene (the level ground on the other side of the river) with the transforming power of God by which 'all things are becoming smooth for His chosen ones'.

It would be possible to give a more detailed analysis of these themes, but that would be the goal of a commentary. Since this section is concerned with developing a method of theme detection, in order to understand the social location of, and the relationship between, author and reader, detailed comment is not appropriate at this stage. However, detailed comment on a specific theme will follow in due course.
We must now address the question, 'Did the method of theme detection produce satisfactory results?' This section began with a search for a method that would perform the same function as sentence diagramming. Ideally, the method should detect themes such as 'God is angry with you' and hence show the importance of references V1.3.1 and V1.3.2 in the structure of Vision 1. The method should be quicker than the laborious task of sentence diagramming, and be able to be performed in an unsupervised manner using a computer. When a simplified lexical mapping (SLM) method was applied to Vision 1, it did manage to show the importance of V1.3.1 and V1.3.2, but it also aided finding a structure and other important themes in the narrative. However, the SLM did not highlight the 'God is angry with you' theme because two out of the three uses of the word 'angry' are found in one reference, and, the bond between V1.1.6 and V1.3.1 is weak compared to other bonds. A conscious decision was made at the beginning of the process to use text reference as the unit of study, rather than the sentence, for ease of interpretation. A loss in sensitivity is compensated for by the quantity of useful information that SLM provides. The quality of that information has yet to be decided.

Narrative Appraisal

This section deals with the question, 'What can be learned about the writer and his readers from this story?' The question is answered by examining the writer's opinion in the subject matter of the narrative. These opinions are signalled in the text by the use of evaluative lexis. Martin classifies four broad areas of interest:

1. **Affect**, which is the resource deployed for construing emotional responses.
2. **Judgement**, which construes moral evaluations of behaviour.
3. **Appreciation**, which construes aesthetic qualities.
4. **Engagement**, which construes commitment to what is being said.

The writer will bring one of these categories to the foreground of the narrative, depending on the mood of the narrative at that point. We will consider each category in turn, using *Vision 1* as the text of interest, but the categories of 'judgement' and 'engagement' will be central to our concern.

---

Affect construes the feelings expressed by the writer as either positive or negative evaluations of a situation. Feelings of happiness (positive evaluation) may be expressed with words such as μακάριος, ἵλαρος, χαρά. Outward expressions of happiness will include words such as γέλαω. Feelings of unhappiness (negative evaluation) may be expressed with words such as:

λυπέω, λυπηρός, κατήφεια, κατηφίς, στυγνάω, στυγνός.  

Outward expressions of unhappiness will include words such as κλαίω.

The writer uses the affect of happiness/unhappiness to describe Hermas' mental attitude to sinfulness. Initially Hermas is happy in his ignorance, "I should be happy if I had a wife of such beauty and character" (VI.1.2). When informed that his thought was in fact sinful, and that God was angry with him, Hermas' mood changes. He becomes filled with grief, the outward expression of which is manifest in his weeping (VI.2.2) Hermas' grief is also displayed in his mental and physical appearance, as with the second Visionary Character asking, "Why are you gloomy Hermas? You who are patient and good-tempered, who are always laughing, why are you downcast in appearance and not merry?" Following the reading of the book (especially the words recorded in the central verse, (VI.3.4), Hermas is asked, "Did my reading please you?" and he replies, "Lady, this last part pleases me". The second Visionary Character then "went away cheerfully". The reader will eventually be informed that the second Visionary Character is the church (VII.2.4.1).

For the writer then, sinfulness brings unhappiness, whereas, happiness is to be found through the mediation of the church and the reading of its book(s). This is a bold statement and perhaps implies a move to a more institutional understanding of the role of the church and its ministry in the community to which Hermas belongs. We will need to return to this concept in the next few chapters.

The writer underlines the role of the second Vision Character, using affect in a number of other ways. This includes feelings that involve intention with respect to an external stimulus, and not merely reaction. Thus, desire (ἐπιθυμία) is used in the negative sense, as a desire for something forbidden, throughout Vision I. It is further described in the text as, 'evil desire' or, 'the desire of wickedness'. The second Vision

---

121 No claims are being made at this stage concerning the ministry in the church to which the writer of The Shepherd belongs. Some argue that there is evidence for a move from, even a repudiation of, a charismatic ministry in The Shepherd, see Lane, W. L. "Social Perspectives on Roman Christianity During the Formative Years from Nero to Nerva". in Donfried, K.P. & Richardson, P (eds.) Judaism and Christianity in First Century Rome.
Character regards such desire as a corrupting influence (V13.1) that destroys simplicity and innocence (V1.2.4). It must be overcome by the exercise of self-control (ἐγκράτεια), the 'daily righteous word' (V1.3.2), and attention to the word read by the church.

We have already had cause to mention the anger of God, the heavens being shut, giving rise to the affect of insecurity. This insecurity manifests itself in Hermas' 'shuddering' and 'grief' (V1.2.1), and are only quelled by the reassurances given by the second Vision Character.

2. **Judgement** relates to moral evaluation of behaviour and involves social esteem, veracity and propriety. Both Vision Characters are accorded a high veracity. After some initial questions (V1.1.7), the veracity of the speeches made by the two women is assumed. Their social esteem is granted because of the authority each is given, but this authority is not because of equal status. This is seen in the fact that the speech of the Vision Character, Rhoda, is 18% of the total text of *Vision 1*, while that of the Vision Character, the Old Woman, is 39%. Hermas' part is small in comparison, only 11%, most of which is questions put to the Vision Characters. 32% of the text is taken up with narration. It also becomes apparent (as mentioned above) that the Old Woman corrects, and adds to, the information provided by the Rhoda character. In terms of speech content and volume, the Old Woman is the leading actor in this vision. Her authority and importance is further emphasised by the fact that she has attendants to remove the chair (V1.4.1) and to escort her from view (V1.4.3). The reading of the book is another display of authority. The one who has the book and reads the words commands the attention of the audience.

Hermas' authority derives from the fact that he has received the vision, which he now communicates as narrator of the story. In the story, Hermas has authority over his family, but he is accused of not exercising this authority in a way that kerbs their waywardness.

For Hermas, then, authority is God-given and benevolent. Those who have authority are to use it to correct subordinates and to encourage righteousness. Subordinates are expected to follow the instruction of those having authority, but not without question; Hermas questioned the accusations of the first Visionary Character. However, Hermas does not question the sayings of the Old Woman since she is not the one accusing him. Rather, the Old Woman acts as an arbitrator between Rhoda and Hermas in the initial exchanges.
2. **Appreciation.** By virtue of the care the narrator takes to set the scene and mood, the point of the story must be important to him. It displays a 'high' view of the holiness/greatness of God and of sin. This view of God is expressed by both Visionary Characters in their speeches (V1.1.6, V1.3.4). The narrator has a strict moral ethic, and, he tries to convince his audience to similar views. Hermas' sin is a result of moral laxity, the effects of which cascades downwards to his family who have become corrupted by 'the things of daily life' (V1.3.1). The writer is therefore concerned by the declining moral standards of the Christian community at Rome.

4. **Engagement** involves the ability of the text to produce solidarity between the characters in the text, and by extension, to produce solidarity between the writer and the reader of the text. When feelings (affect), information (judgement), or quality of life values (appreciation) are shared with another person, they can evoke bonding, alienation or neutral responses. Since Vision 1 foregrounds the use of affect, the writer wishes to achieve solidarity with the reader based on the assumption of shared emotional experience.

Hermas' sin is exposed by the Vision Character (Rhoda) V1.1.5 in a very direct manner. Hermas responds defensively. The initial stage of this encounter involves alienation between the two characters. The reader can form a bond of empathy with the characters either as the person wronged (Rhoda) or the person exposed (Hermas); these are common human experiences. In order to stop the alienation from developing into a potentially complete rift the writer relieves the tension with the affect behaviour, "She **laughed** and said to me". The seriousness of the charge is not diminished, as the rest of the Vision Character's speech shows, but the relationship between the characters is somewhat recovered. This skilfully constructed dialogue reinforces the writer's message that sin results in alienation between sinner and God (Heavens were shut) and society.

The grief that Hermas feels enables the encounter with the second Vision Character to begin in a neutral manner. Although the accusations are repeated Hermas does not respond defensively. By the close of the encounter, there is a bonding between the characters. This bond is created/cemented by the reading aloud of the book. The writer engages the reader in the story using the psychology of human inquisitiveness, 'what is written in the book?', we know the first Vision Character is Rhoda, who is the second Vision Character? The reader is left guessing and waiting to know the answer. Hermas' engagement with the words of the book
changes his mood and removes feelings of alienation. The first Vision closes with hope; Hermas is not left in isolation and alienation.

It has been shown in the analysis above that it is possible to perform an analysis based solely on the grammatical construction of the text. This analysis is able to provide similar results to a sentence diagram, which takes a great deal of time to do, and is able to provide clues to the relationship that exists between writer and reader. A great deal of information can be discovered, which, as noted previously, is suitable for a commentary, but does not suit the purpose of this thesis. The method that has provided such good results is next applied to the complete text of The Shepherd.

**Application of Method to The Shepherd of Hermas.**

We are concerned to find themes that occur in The Shepherd as a whole, and not themes which may be specific to sections alone (eg. as Vision 1 above). The classes of words chosen for the analysis are the same as described previously, but the text is examined by section rather than verse, so, Vision1.1, Vision1.2 etc., NOT Vision1.1.1, Vision1.1.2. To examine by verse would require a cross-tabulation too large for the software to handle. Consequently we are now limited by software versatility.

Having obtained the matrix of Simple Lexical Links, an adjustment was made to the sum of links in order to standardize for section length. It would be expected that a large section of say, 1000 words, has the potential to form more links than a section which might only contain 100 words. Consequently, if each of the two sections had 5 links with other sections of text, it would be structurally more revealing for the 100 word section than the 1000 word section.

Next, the problem is to find a way to describe and graphically represent the network of textual linkage. Because the number of links increases with decreasing bond strength (see graph) the network can become very complex very quickly. The problem must therefore be tackled piecemeal.
The frequency of the links in the data is shown below:

If we look at very strong links first, and take anything \( \geq 12.0 \) as being a very

186
strong link/bond, we can construct a simplified network. This network is shown on the following page.

The first thing to notice in the network displayed below, is the very prominent function of Mandate 1. It links with several other sections of the Mandates and with three sections from the Similitudes. Mandate 1 is clearly a theme originator of central importance. You will also remember from chapter 1 that M1.1.1 is the section of *The Shepherd* most quoted by the early church Fathers. This cannot be a coincidence.

Next we can observe the several fragments of themes, mainly comprising sections from the Similitudes and Visions. The boxes are colour coded to aid the visual appearance of the diagram. The colours are:

- Green boxes represent theme originator.
- Blue boxes represent theme terminator.
- Yellow box a theme carrier/transformer.

Before looking at the links in order to discover themes, we can note that Mandate 1, which is only a very short text, has strong links to most other Mandates, notably Mandates 2, 3, 7, 8, 9, 11. These Mandates are single chapter Mandates. Mandate 6 consists of two chapters, and it can be seen that Mandate 1 links with the second chapter of that Mandate. Mandates 4, 5, 10, 12 are noted for their absence, having low bond strengths to Mandate 1, and, they are multi-chaptered Mandates.
Where Mandates 4 and 12 are found in the network, they are linked only
through another Mandate or Similitude. None of the Visions is found in the network of strong links, rather the stronger bonded Visions appear as fragments. Similarly there are four Similitude fragments. Because of this loose association (lexical bonding) of the Visions, it can now be seen why the Visions might be circulated in the early church detached from the remainder of The Shepherd.

This network has some interesting similarities to Coleborne’s multi-author theory. He suggested that Mandates 1 - 12.3.3 were written first, Similitudes 1 - 7 were then added by another, later, writer. While not in complete agreement, the lexical bonding net shows a similar pattern with Similitudes 5, 6, 7 being bonded to Mandate 1. It is now possible to see Coleborne’s error.

In his research Coleborne had used ALL grammatical categories in his analysis. Using index cards (because the work was done before computing was cheap and easy) he constructed his ‘Apparatus Discernendi’, carefully noting attributive and predicative adjectives etc. Inadvertently, Coleborne had discovered the thematic lexical bonding similar to that utilized in this thesis. However, Coleborne then assumed that his data had distinguished several authors, but what he had actually found was several different major themes running through the text. It now becomes clear that ‘descriptor’ grammatical categories can be used to find themes, whereas ‘function words’ must be used to find authors.

The importance of Mandate 1 lies in its formulaic/creedal nature. Osiek’s also points out that it is the only Mandate to open with no form of address. These two facts tend to suggest the importance of this opening Mandate as the starting point for an explanation of what it means in practice ‘to live to God’. This approach, beginning with an important assertion (text) and developing a wide-ranging theological and practical understanding is reminiscent of Jewish paraenesis. Indeed, the Mandates contain a number of concepts and ideas that seem to be of Jewish-Christian background. The Mandates, therefore, suggest that Hermas’ group had a Jewish-Christian background.

The bond diagram above has served a purpose in providing limited, but important, information. However it is not very interesting because it only gives limited information about major themes present in The Shepherd, simply because M1.1 dominates the diagram due to the common terms ‘fear’, ‘righteousness’, ‘wickedness’, ‘live to God’, ‘believe’, ‘keep this commandment’, which are also

122 Osiek, C. The Shepherd of Hermas. p.103
common to the majority of Mandates. While these are important major themes, the large bond strength removes focus from other interesting features, therefore, the bond strengths were re-evaluated ignoring the effect of M1.1. That does not mean that M1.1 is ignored, we have already noted its importance, rather its removal allows a more complicated structure to be found in the text.
The diagram shows theme initiators in green. In terms of the number of bonds formed, the most important is M4.2 (5 bonds) followed by M12.3 (2 bonds), S5.1 (2 bonds), V2.1 (2 bonds) and S6.4 (2 bonds).

Theme terminators are shown in orange, where S7.1 (6 bonds), S8.1 (3 bonds), S5.3 (2 bonds) are seen as central elements of the network of bonds.

A large network of bonds can be seen, along with a number of fragments. In trying to identify the themes shown in the diagram it is best to begin with the largest theme initiator M4.2, which is bonded to theme terminators M8.1, M11.1, S5.3, S7.1, S8.6. The diagram below shows this group and the call-out boxes show the important link-words that form the bond.
There are some words common to all the bonds (*κύριε* and *κυρίον*) and some words that are characteristic of some bonds only. We can see that M4.2 bonds with M11.1 by the word *ἀφρων*. Therefore there is a link around the theme of ‘foolishness’. Proceeding in this manner we can find themes concerning ‘doing good and refraining from evil’ (M4.2 – M8.1), ‘Repentance’ (M4.2 – S8.6), ‘Keeping the commandments purifies the heart and fills the soul’ (M4.2 – S5.3), ‘Repent with heart and soul’ (M4.2 – S7.1). It can be seen that the links to S8.6 and S7.1 both deal with the theme of repentance but in two different ways, which could be summarized as ‘repentance required’ and ‘repentance given’.

Performing a similar exercise for all bonded groups in the diagram we can obtain the following table of themes:

<table>
<thead>
<tr>
<th>Bond</th>
<th>Theme</th>
</tr>
</thead>
<tbody>
<tr>
<td>V2.1 - V3.1</td>
<td>Hermas at prayer.</td>
</tr>
<tr>
<td>V2.1 - V7.1</td>
<td>Hermas thought worthy.</td>
</tr>
<tr>
<td>V2.4 - S7.1</td>
<td>Hermas and his house.</td>
</tr>
<tr>
<td>M12.3 - S7.1</td>
<td>The glorious commandments; repentance of Hermas’ house</td>
</tr>
<tr>
<td>M12.3 - S6.1</td>
<td>The glorious commandments and good desire; faith and virtue</td>
</tr>
<tr>
<td>M4.2 - S5.3</td>
<td>Keeping the commandments is good for the soul.</td>
</tr>
<tr>
<td>V3.4 - V3.8</td>
<td>The building of the tower: men and maidens.</td>
</tr>
<tr>
<td>S9.3 - S9.4</td>
<td>Stones from the deep, men and maidens.</td>
</tr>
<tr>
<td>M4.2 - S7.1</td>
<td>Repentance.</td>
</tr>
<tr>
<td>M4.3 - S8.6</td>
<td>Repentance.</td>
</tr>
<tr>
<td>M4.2 - S8.6</td>
<td>Repentance.</td>
</tr>
<tr>
<td>M4.2 - M8.1</td>
<td>Good and evil deeds.</td>
</tr>
<tr>
<td>M7.1 - M8.1</td>
<td>Evil deeds overcome by right use of fear.</td>
</tr>
<tr>
<td>S5.1 - S7.1</td>
<td>Clean hearts and evil deeds.</td>
</tr>
<tr>
<td>S5.4 - S5.5</td>
<td>Parable of vineyard.</td>
</tr>
<tr>
<td>S9.27 - S9.28</td>
<td>Image of trees and fruitfulness.</td>
</tr>
<tr>
<td>M4.2 - M11.1</td>
<td>Foolishness.</td>
</tr>
<tr>
<td>S6.4 - S9.14</td>
<td>Foolishness.</td>
</tr>
<tr>
<td>S6.4 - S6.5</td>
<td>Foolishness of luxury and deceit.</td>
</tr>
<tr>
<td>S9.10 - S9.11</td>
<td>Cheerfulness.</td>
</tr>
<tr>
<td>S3.1 - S4.1</td>
<td>Withered trees in winter.</td>
</tr>
<tr>
<td>S5.1 - S5.3</td>
<td>Acceptable fast.</td>
</tr>
</tbody>
</table>

Themes that are similar are given the same colour-coding, for example themes which seem to be about Hermas are in pink. Some themes are overlapping, such as the commandments when specifically mentioned with respect to Hermas’ house. There is no order of importance implied in the position of the theme in the table, the themes are arranged randomly, except where they are in the same colour-coded
What the Themes Tell Us About Hermas And His Community.

Before we begin this section there are certain difficulties that have to be faced. Wilson\textsuperscript{123} accepted single authorship, and used the text of *The Shepherd* as a source of biographical information about Hermas. Dibelius\textsuperscript{124}, however, in his extensive commentary had concluded that the whole work was allegorical, including all of the autobiographical sections present in *The Shepherd*. Most commentators tend to agree with Wilson that parts of Hermas' life and world-view can be discovered from a reading of his work, indeed, Lampe\textsuperscript{125} gives several pages to a discussion about Hermas' level of education, literary ability and language. His conclusion is that Hermas is "a man from the common folk", and is therefore a typical representative of Roman Christianity. Audet\textsuperscript{126} concludes from his studies that Hermas was born of Jewish parents, who had belonged to the Qumran community. They were taken as slaves to Italy where their son Hermas was also sold into slavery. Audet has his supporters. This thesis does not join with Audet's views, but assumes that autobiographical material found in *The Shepherd* does give an insight into Hermas the writer.

Bearing this in mind, working down the list of themes given above, we will be able to identify some of the major issues that occupied the mind of Hermas. First on the list is a series of autobiographical issues:

| V2.1 - V3.1 | Hermas at prayer. |
| V2.1 - V7.1 | Hermas thought worthy |
| V2.4 - S7.1 | Hermas and his house. |

Hermas is concerned to show his authority as a receiver of visions and as the human agent through whom God speaks to the church. It was while praying (V1.1.3-4) that the vision was given to him, particularly while confessing his sin in this first vision. 'The Church' convinces Hermas that the healing of his own sin, the sins of his family and the church would be aided by his continuance in prayer (V1.1.9). Piety is important for the healing/health or purity of the church. The

\textsuperscript{123} Wilson, W.J. "The career of the Prophet Hermas". *HTR* (1927) 20, p.21-62.
\textsuperscript{124} Dibelius, M. *Der Hirt des Hermas*.
\textsuperscript{125} Lampe, P. *From Paul to Valentinus*, pp.227-236.
\textsuperscript{126} Audet, J.P. "Affinités littéraires et doctrinales du Manuel de Discipline". R.B. (1953) 60, pp.41-82.

second vision occurs while Hermas is praying also (V2.1.2). This time his prayer is not confession but praise, and about the fact that God had thought him **worthy** of revelation. The word ‘worthy’ is a very important appraisal word for Hermas’ authority. His authority to communicate the message to the church is legitimated in God accounting him worthy. In his third vision Hermas is again found in prayer, confessing his sin (V3.1.6) before being interrupted by The Church and six young men. It was while praying at home that Hermas has his visionary encounter with The Shepherd. There is then a long gap in the narrative until S9.11.7 where mention is made of Hermas’ stay with the maidens who did nothing else but pray. Hermas also prayed unceasingly.

The visions, then, are proofs of Hermas’ piety, but they also tell us more. Hermas receives his revelations when he is alone\(^{127}\). He is commanded to write, or tell, the visions to the church community after they have occurred. These visions are not part of the corporate worship of the church. Since the visions are subsequently recorded and reported to the church, they might contain traces of their transmission to the church, that is, they might well contain marks of an oral re-telling of the experience. This may indeed account for the sudden shift of the personal pronoun in some parts of the text from singular to plural. Vision 3.8.9-10 is an example of such a shift:

\[
\text{άρκετή σοι ἡ υπόμνησις αὕτη καὶ τῶν ἁγίων καὶ ἡνακαίνωσις τῶν πνευμάτων ὑμῶν ἄλλῳ ὑπὸ σοὶ μόνῳ ταῦτα ἀπεκαλύφθη ἄλλῳ ἵνα πᾶσων διήλωσις αὐτά.}
\]

The change to you (pl) might be a reminiscence of oral delivery to the church. Hermas thus declares his authority as receiver and transmitter of the revelation.

Next we note the role of the commandments:

| M12.3 - S7.1 | The glorious commandments repentance of Hermas' house. |
| M12.3 - S6.1 | The glorious commandments and good desire: faith and virtue. |
| M4.2 - S5.3 | Keeping the commandments good for the soul. |

For Hermas the commandments are a set of rules by which **all** Christians of the church should conduct their lives, but even he notes that they are ‘very hard’ to keep.

---

127 Reiling, J. *Hermas and Christian Prophecy*. p.163

195
Without going into the detail of the commandments, we can see that they are set out in a typical paraenetical form of question and answer. This form of teaching may well be based on the Jewish system of paraenesis\textsuperscript{128} and seems to be but one aspect of a wider Jewish-Christian background to \textit{The Shepherd}. Hermas also borrows material from the Hellenistic world, and it is this mixture that Reiling\textsuperscript{129} finds surprising.

Hermas views the changing church with despair because he identifies a lax morality that indicates the church is accepting the standards of Roman society. He writes about the abuse of wealth by some within the church, the greed and self-serving of some of the leaders of the church, he comments on the continuing quarrelling, and the relaxation of ethics in many areas of church life. Hermas is so concerned about the blurring of community boundaries that he feels some sort of 'rule' is required. The \textit{Mandates} set out the rule of moral and spiritual ethics that the community should keep.

In Mandate\textsuperscript{12.3.4} Hermas describes the commandments as being 'great and glorious'. Of the 43 times the adjective 'glorious' is used in \textit{The Shepherd}, only twice does it relate to the earthly idea of honour. Its most common use is in the description of the 'otherworldly' appearance of vision figures. Hermas thus portrays these commandments as being of spiritual origin, and therefore authoritative. This tells us a great deal about the Hermas community.

Maier\textsuperscript{130}, relying heavily on the work of Geertz\textsuperscript{131} and Wilson\textsuperscript{132}, discusses at some length the issues involved:

The importance of Hermas' paraenesis is that it is designed to shape the behaviour of members who are participants in the transcendent reality of the church... The use of injunction and the establishment of regulations, such as those limiting the amount of business one may involve oneself in, are important means of ensuring that the sect maintains a distinctive identity separate from the outside world.

The Hermas community can therefore be seen as a community at odds with pagan society, yes, but feeling very vulnerable to its influence in the church. The Hermas community is also uncomfortable with the failure of the catholic church in

\textsuperscript{128} Snyder, G. "The Shepherd of Hermas". in Grant, R.M. (ed) \textit{Apostolic Fathers}.
\textsuperscript{130} Maier, H.O. \textit{The Social Setting of the Ministry}. p.76.
\textsuperscript{131} Geertz, C. "Ethos, World-View and the Analysis of Sacred Symbols". \textit{Antioch Review} (1958) 17, pp.421-437.
\textsuperscript{132} Wilson, B.R. \textit{Patterns of Sectarianism}.
Rome to halt the decline, consequently they strengthen their boundaries with the rule of the commandments. In doing this, the community becomes inward looking and is in danger of becoming isolated.

Next in the list of major themes:

| V3.4 - V3.8 | The building of the tower: men and maidens. |
| S9.3 - S9.4 | Stones from the deep, men and maidens.     |

The vision of the Tower is an important part of The Shepherd and occurs twice in the work, Vision 3 and Similitude 9. The attendants of the visionary figure of The Church play an important role in the vision. Osiek offers an explanation of their role in the story:

The six primary angels are the ones who previously escorted the woman church and carried her chair (V1.4.1), and are generally understood to be those first created and responsible for cosmic supervision in Jewish apocalypticism (1 Enoch 90:20-22) – except that they are usually seven, not six.

While this is instructive, the importance of the Tower vision lies in its use of analogy regarding the stones. Hermas is able to use the stone to represent various Christian characteristics. Only the perfect, square stones fit directly into the Tower. These are the Apostles and leaders of the church. Other stones are explained that will need attention from the builders before they can take a full part in the construction of the Tower. He uses, for example, the image of a round stone for those who have faith and wealth. The round stone would destabilise the Tower so would require some stone-masonry expertise to get them fit for the building. Hermas is then able to describe how the ‘correct’ use of wealth and repentance for their greed will first need attention for them to fit into the Tower. All of this is another way of exercising social control over the Hermas community, and reinforces all that was said in the previous section above.

Next,


There are 22 references to truth in The Shepherd, many of them in connection with business, wealth and luxury. For Hermas truth is not just the absence of a lie,

133 Osiek, C. The Shepherd of Hermas. p.69.
truth is a way of being. The spirit that God has placed within is a spirit of truthfulness because God is true (M3.1.1). In order that this spirit dwells happily within the Christian, the Christian must do good. The departure of the spirit brings misery to the person who does evil. In M3.1.2 Hermas likens those who do not follow truth as ἀποστερητῆς (robbers). He seems to use a business or banking idea saying that those who lie have robbed God of his deposit. For Hermas, then, God has invested His truth in His people, to deny that truth is equivalent to robbing God of His investment. There is a great deal of affect and judgement in the narrative of Mandate 3. The judgement (moral stance) appears in M3.1.2 and includes words such as ‘robber’, ‘defraud’, ‘not restore’ and ‘defiled’. The affect (emotional response) appears in M3.1.3 and includes words such as ‘weep’, ‘weeping’.

In M10.1.5 Hermas states:

άλλα καὶ ὅταν ἀκούσωσι περί θεότητος καὶ ἀληθείας ὅ νοῦς αὐτῶν περί τὴν πράξιν αὐτῶν καταγίνεται καὶ οὐδὲν ὅλως νοοῦσιν.

Those who do not understand are those who do not listen to the truth. Hermas describes these people as having ‘only faith’ and who become mixed up with business, riches, heathen friendships and the occupations of the world (M10.1.4). Those who have only faith must therefore be those who lack a deeper understanding of the things of God, because, like the seed sown amongst the thorns, they are choked by the cares of this world. Luxury and deceit pervert the servants of God from the truth (S6.2.1). Finally,

ἐάν ταύτα φιλάξης καὶ ἀπό τοῦ νῦν πᾶσαν ἀληθειανκαλλήσης δυνάσῃ σεαυτῷ ζῷν περιποίησασθαι Καὶ ὃς ἄν ἀκούσῃ τὴν ἐνειλήν ταύτην καὶ ἀπέξεται τοῦ πονηροτάτου ψεύσματος ζήσεται τῷ θεῷ. (M3.1.5)

Hermas is therefore concerned about a laxity of religious discipline which is allowing worldly attitudes to creep into the church. He considers it his duty to warn the church about these dangers.

The next theme to consider is:

| M4.2 - S7.1 | Repentance. |
| M4.3 - S8.6 | Repentance. |
Most commentators are agreed that this is the major theme of *The Shepherd*. Pernveden, however, regarded repentance as an prominent theme in *The Shepherd* while not being convinced that it was central.

The difficulties that scholarship has had both in arriving at an uncontradictory understanding of the very problem of metanoia, and in arriving at a consistently systematic interpretation of *The Shepherd* on the basis of the concept of metanoia makes us suspect that the idea of metanoia is not the real key to the complex of problems in *The Shepherd*. **The task, then, must be to look for another approach that is more practicable.** This must consist of a theme that is both central and characteristic of the whole text and is also of such a nature that it can include the problem of metanoia as well and let it be fitted into a natural and organic context\(^\text{134}\). (Bold mine)

That which Pernveden sought is partially satisfied by the methodology of this chapter. Admittedly, Pernveden resisted the way that some scholars used *The Shepherd* to help support a doctrine of penance, consequently he turns his attention to the concept of ‘church’ in *The Shepherd*. He therefore makes the theme of the church central to the text, repentance then being a prominent part of that theme. The analysis of this chapter suggests that repentance is one of a number of important themes. From the analysis suggested in this chapter of the thesis it can be seen that word repetition focuses on M4.2 and M4.3 are the theme initiator sections regarding repentance. The theme runs through the text leading to some major conclusions in S7.1 and S7.6. It is important to re-iterate that the analysis was used to find **major** themes, hence the concentration on a few sections of the text. In fact, there are 95 words relating to ‘repentance’ in *The Shepherd* which means that nearly every section of the work includes a reference to the subject. The theme is always in the background of the narrative, but it surfaces particularly in the sections noted in the table above.

In M4.2.2 repentance is linked to understanding (σῶςως). This understanding is something that comes from God, so in S5.4.4 Hermas is told "wherefore do you not seek understanding from the Lord". All of the uses of the word in *The Shepherd* imply that this understanding is enlightenment, so too in M4.2.2. Repentance is enlightenment because God causes the sinner to see he has

---

\(^{134}\) Pernveden, L. *The Concept of Church in The Shepherd of Hermas*. p.11,12.
done wickedly, hence he repents, and then he does good abundantly, and humbles his soul. However, Hermas is concerned about the lax practices creeping into the church and he was not the only one. There were some in the church who thought it possible “both to defer and to repeat postbaptismal repentance”\textsuperscript{135}. Perhaps such ideas were brought to the attention of Paul when he writes:

\begin{quote}
\textit{oτίνες ἀπεθάνομεν τῇ ἁμαρτίᾳ, πῶς ἔτι ζήσομεν ἐν αὐτῇ;} (Rom 6:2)
\end{quote}

If we accept Bruce’s\textsuperscript{136} idea that the New Testament letter to the Hebrews is a document of Roman Christianity, we also note that there was a strict group in Rome who denied a second repentance. There might be an allusion to this group in M4.3.2, which may refer to Heb 6:4, but this is not certain:

\begin{quote}
ἐὰν δὲ ὑπὸ χείρα ἁμαρτάνῃ καὶ μετανοήσῃ ἡ σύμφορον ἑστίν τῷ ἀνθρώπῳ τῷ τοιούτῳ δυσκόλῳ γάρ ζήσεται
(M4.3.6)
\end{quote}

Hermas takes a middle stance, suggesting that there is a second chance (M4.3.6), one new possibility of repentance after baptism. By allowing this, Hermas gives hope to those who have sinned, but he also ensures that the responsibility of the penitent is to resist sin,

\begin{quote}
τῶν οὖν μετανοοῦντων εὐθὺς δοκεῖς τὰς ἁμαρτίας αφίεσθαι Ραντελῶς ἀλλὰ δεῖ τῶν μετανοοῦντας βασανίσαι τὴν ἑαυτοῦ ψυχὴν καὶ ταπεινοφρονῆσαι ἐν πάσῃ πράξει αὐτοῦ ἱσχυρῶς
\end{quote}

\textsuperscript{135} Lampe, P. \textit{From Paul to Valentinus}, p.94.
καὶ θλίβηναι ἐν πάσαις θύψεσι πουκίλαις.
(S7.1.4)

This comes close to the idea of penitence with the associated components of humbling and affliction, but for Hermas it means being pure in heart. We see this in S7.1.6

μόνον παράμεινον ταπείνοφρονῶν καὶ λειτουργῶν τῷ κυρίῳ πάση καθαρᾷ καρδίᾳ ... καὶ πορεύον ἐν ταῖς ἐντολαῖς ... αἱ δυνάμεις σου ἡ μετάνοια ἴσχυρὰ καὶ καθαρὰ εἰναι.
(S7.1.6)

This again becomes apparent in S8.6.2-6, where the parable of the sticks is explained to Hermas. Repentance is given to those whose heart was pure. He then gives a list of those to repentance was not given. Those for whom there was no repentance include apostates, betrayers of the church, blasphemers and those ashamed of the name of the Lord, hypocrites and false teachers have hope of repentance, but if they refuse to repent they will lose their lives.

οὗτοι οὖν εἰς τέλος ἀπώλειτο τῷ θεῷ βλέπεις.

For the Hermas community then, we see that repentance and purity are important issues. These concerns are again an attempt to keep the church free of worldly morality and ethics.

| M4.2 - M8.1 | Good and evil deeds. |
| M7.1 - M8.1 | Evil deeds overcome by right use of fear. |
| S5.1 - S7.1 | Clean hearts and evil deeds. |

This theme is related to the theme of repentance. The purity of the Hermas community is maintained by doing good. M8.1.3-6 lists the evil deeds that a servant of God must not do and M8.1.9-11 lists the good deeds a servant of God must do. In order to ensure the doing of good deeds, M7.1.4 shows the importance of 'fear the Lord'. This right use of fear is stated succinctly as:

ἐάν γάρ θέλης τὸ πονηρὸν ἐργάσασθαι φοβοῦ τὸν κύριον καὶ οὐκ ἐργάσῃ αὐτῷ; ἐάν δὲ θέλης πάλιν τὸ ἄγαθὸν ἐργάσασθαι φοβοῦ τὸν κύριον καὶ ἐργάσῃ αὐτῷ.
The sections from the *Similitudes* link the doing of good with purity of heart, so:

μηδὲν ποιησάσθη ἐν τῇ ζωῇ σου ἀλλὰ δουλεύσον τῷ κυρίῳ ἐν καθαρᾷ καρδίᾳ.
(S5.1.5)

<table>
<thead>
<tr>
<th>S5.4 - S5.5</th>
<th>Parable of vineyard.</th>
</tr>
</thead>
<tbody>
<tr>
<td>S9.27 - S9.28</td>
<td>Image of trees and fruitfulness</td>
</tr>
</tbody>
</table>

These two sections are not linked by any bonding in the text, but I have placed them together for discussion.

S5.4 bonds to S5.5, this might seem an obvious bonding since the sections are part of the explanation of the parable of the vineyard. The links include the words ‘field’, ‘vineyard’, ‘parable’, ‘master’, ‘servant’, ‘Lord’, but the other important words are αὐθάδης (self-willed, wilful, dogged, stubborn, contumacious, presumptuous) and γνωστὸς (known). The important verses of these sections then are S5.4.2&3 and S5.5.1:

ό δὲ ἀποκριθεὶς μου εἶπεν αὐθάδης εἰ λιαν εἰς τὸ ἐπερωτάν ... ὅσ τὸν δούλον ήτοι θεοῦ καὶ ἐχὴ τὸν κυρίον ἐγενέσθαι παρ’ αὐτῷ σύνεσιν καὶ λαμβάνει καὶ πάσαν παραβολὴν ἐπιλύει καὶ γνωστὰ αὐτῷ γίνονται τὰ ῥήματα τοῦ κυρίου τὰ λεγόμενα διὰ παραβολῶν
(S5.4.2&3)

εἶπον όσι φησὶ καὶ ἀρτι ὅτι πανούγρος εἰ καὶ αὐθάδης ἐπερωτῶν τᾶς ἐπιλύεις τῶν παραβολῶν ἐπειδή δὲ οὕτω παράμονος εἰ ἐπιλύσω σοι τὴν παραβολὴν τοῦ ἀγροῦ καὶ τῶν λοιπῶν τῶν ἄκολουθων πάντων ἵνα γνωστὰ πάσιν πουήσῃς αὐτά ἄκουε νῦν φησίκαλ σύνε αὐτά.
(S5.5.1)

Because of Hermas’ persistence in asking for explanations, the shepherd tells him that, God makes known the meaning of the parables to His servants, but Hermas is then told that he must make them known to everyone. Hermas is stating again the authority that has been given to him to speak on God’s behalf.

202
The other two sections, S9.27 and S9.28 are not related to those discussed above. The emphasis in these two sections is on the trees. S9.27 is the vision of the tenth mountain where the sheep are sheltered by the trees. These trees are the leaders of the church who have acted with compassion in caring for the poor. In S9.28, the vision of the eleventh mountain, the trees are fruit trees. These are identified as those who have suffered for the name of the Son of God.

| M4.2 - M11.1 | Foolishness. |
| S6.4 - S9.14 | Foolishness. |
| S6.4 - S6.5 | Foolishness of luxury and deceit. |

Hermas defines foolishness (ἄφρων) as a lack of understanding M4.2.1. As we have seen above, understanding occurs when God enlightens the sinner, especially with respect to his sin in the case of repentance. Foolishness, then, is the opposite of understanding, there is no enlightenment and the spirit of God, who gives understanding, is absent. This foolishness manifests itself in the words of the false prophet and those who enquire of them (M11.1). Hermas, in Mandate 11, describes the false prophet as being empty (κεκοιμημένος), that is, devoid of the spirit of God. If the false prophet is devoid of the spirit of God, the words that he speaks cannot be words of understanding, they are words of foolishness (M11.1.11).

Hermas notes that the false prophet does not enter the assembly of the righteous, but speaks in corners and only to those who are likewise ‘empty’ (M11.1.13). Therefore, those who seek the false prophet are foolish (M11.1.4).

In the complex of sections S6.4, S6.5 and S9.14, foolishness is also the theme, although there are no bonds between the Mandates and Similitude sections. In S6.4&5, the consequences of foolishness are shown in practical examples. Hermas concentrates on the foolishness of luxury (τρωφή). Those who live in luxury and forget God will be punished (S6.4.2) and Hermas is told that he is foolish and does not understand the power of punishment. Those who live in luxury are described as foolish (S6.5.3), presumably because they forget God and the spirit of God is not in them. However, there are those who have luxuries but do not forget God. Hermas therefore introduces a ‘two-ways’ philosophy to luxury. There is a harmful luxury, resulting in punishment, and a profitable luxury, bringing life. The foolish do not understand (not enlightened) concerning the right use of wealth, whereas those who do understand (are enlightened) use wealth ‘correctly’ to bring life.

In S9.14, the related ideas of foolishness and understanding are still present,
but the emphasis is again shifted. Hermas is described as foolish because he does not understand the meaning of the Tower built on the rock. The understanding is given to him that the Son of God supports creation and those who bear His name with their whole heart.

Therefore, for the Hermas community there are two levels of cognition, which can be defined as understanding and foolishness. The believer is to seek understanding. The spirit of God gives this understanding to those who ask. It is almost the possession of a higher knowledge concerning ordinary things, such as, luxury, fear, repentance. The world, and those in the church who do not seek understanding, are ἄφωνοι.

These two sections are difficult to understand because of the nature of the material contained in them. Osiek regards the maidens, into whose custody Hermas is placed, as female personifications of the Greco-Roman virtues. The key to understanding the section is not related to questions of celibacy in male-female relationships, rather, it is ‘cheerfulness’ (λαρός). The cheerfulness of the maidens and of Hermas is due to the completion of the Tower. The Tower is swept and cleaned ready for the visit of the lord (S9.10.3&4). The ‘filling of the marks of the stones’ (S9.10.2) has been a subject of debate. Does it mean the holes which marked the places the stones had been taken from were filled-in, so that everything became smooth (ὅμαλός), or does it refer to the stone walls of the Tower itself (I tend to favour the latter interpretation). Whatever the intention of Hermas, the main theme is that the Tower is completed and this causes great cheerfulness. Hermas experienced cheerfulness too, in praying and δέησις προῦτα κυρίου (S9.11.8).

For Hermas, true joy is found in the service of the church, prayer and reading/teaching.

In this vision Hermas sees trees that are barren of leaves, they all appear to be in the same condition dry/diseased (ἐξηρέτω). Christians in this world are then

137 Osiek, C. The Shepherd of Hermas. p.228.
likened to these trees, like trees in winter it is difficult to tell the good trees (the righteous) from the dead trees (sinners). When the summer comes the good trees blossom and can be distinguished from the dead trees which remain as they were. In the eschatological summer that Hermas foresees, the unrighteous will be burnt up like wood. It is therefore imperative that the Christian is faithful to God now, and that means in business life(S4.1.5). By serving God with a pure mind they will bear fruit for the world to come (S4.1.8).

Hermas calls on his community to continue to act righteously, even if there seems to be little effect now. It implies the community was struggling to come to terms with its role in the world, perhaps some were discouraged because of their humble circumstances. Hermas encourages a looking to eschatological time.

In this section Hermas is found in a period of personal fasting. When asked what he is doing in that place (on a mountain) so early he replies that he has a station (οταγιώνα). This was probably some new practice in the church because the Shepherd is unaware of its meaning. Hermas is then given understanding of what a fast should be, that is to do nothing evil, serve the Lord with a pure heart, keep the commandments, fear God and refrain from wickedness.

Kai tauta ean ergado megalen upestelaioun poiaseis kai dektiyn ti.model. (S5.1.5)

CONCLUSION

A simple lexical pattern analysis has been shown to work well for Greek texts. It allows extraction of themes from written texts that, with the right software, could be done in an automatic unsupervised manner. This method avoids some of the more subjective approaches to analysing texts, and sometimes highlights themes that have not been considered before (eg. foolishness in the above analysis). The tendency of subjective methods is to consider themes/meanings that may be of importance to the social context of the analyst rather than the text, and these themes can sometimes be

---

138 See Osiek, The Shepherd of Hermas. p.169 n.6. Hilhorst suggests that the term arises from a Latin translation of the Hebrew phrase, “stand before the face of the Lord” where the Latin verb for stand appears as station. Snyder suggests the term is of Roman military origin, meaning to stand guard."
taken-up by other analysts, producing a vast literature on the subject. It has thus become impossible, for example, to consider *The Shepherd* without paying service to the concept of 'rich and poor' as first introduced by Osiek. However, the analysis of themes does not find it to be one of the major themes, except where it is included in the major theme of purity. For Hermas then, the relationship between the rich and poor is representative of a much greater idea, that of purity in the community.

This chapter has developed a method which can function at a number of levels. It has been applied here to major themes only, in an effort to discover what occupied the minds of Hermas and his group. By doing so, the results of Coleborne's analysis, which for so long had been dismissed without comment, can now be explained. Following the major themes through the text has also shown that the Mandates and Similitudes form a literary unity, for it would be difficult to conceive of one part without the other without damaging the narrative. The Visions, however, could be (and were) separated from the narrative without a damaging effect.

The level of bonding between sections was chosen, in this analysis, so that the theme map was not too complicated to draw and interpret. It is possible to use graphical software that would do this job automatically too. Another way of dealing with the data is to concentrate on a particular theme. In this way the number of links to form a bond could be decreased so that the theme in its entirety could be mapped through the text. This would then show the main 'strand' of the theme through the text, and the subordinate ways in which the theme is developed.

The methodology employed in this chapter is also an ideal way to begin a commentary upon the text since, it is able to provide information for the intelligent analyst to comment on and expand into other areas of research.

Concentrating on the major themes in this way, we were able to identify some of those things which were of concern to Hermas at the time he wrote *The Shepherd*. These themes suggest that the Hermas community was going through a period of difficulty (crisis would be too emotive a word to use, and perhaps not justified from the evidence). The difficulty is centred on the laxity of the church and its conformity with the pagan world, or to put it another way, Hermas is concerned for the purity of the church. This laxity is found, in Hermas' opinion, throughout the church, including amongst its leaders.

Hermas has a great deal to say about those who put their business concerns before that of righteousness. He does not say that ALL of those with businesses lead to unrighteous, but he is concerned that Roman business practices are at odds with
care for the widow, the orphan and the poor. Wealth means that some in the church live in luxury while others struggle to survive.

It is now possible to answer the remaining questions that were posed in the introduction, namely:

**Can any dislocation of textual narrative flow be detected in *Visions* 1-5 that may indicate misplaced pages of the text?**

The flow of major themes shown on page 178 does not include many major themes from *The Visions*. This question was raised following Kirkland's hypothesis that there had been a displacement of parchment leaves at some time in the ancient past. He argued that V3.2.4 – V3.10.6 is in fact the second appearance of the Woman Church and that V3.1.1 – V3.2.3 should be, along with *Vision* 2, the third appearance of the Woman Church. It is interesting that the theme flow diagram links V2.1 to V3.1, and they, in turn, are linked to the main structure by S7.1. V3.4 is linked to V3.8, on the other hand, as a fragment. This seems to imply that there might be some truth in Kirkland's hypothesis, although we cannot draw any firm conclusions from the current data. In order to investigate this possible dislocation more thoroughly, the methodology used here should be applied to *The Visions* only in order to get more detailed information.

**Is there evidence of dislocation in the text, particularly M4.2.1 - M4.3.7?**

Osiek reported the possible dislocation of Mandates 4.2 and 4.3, which should, in fact, be part of Vision 5. This question cannot be answered on the basis of the results obtained here. M4.2 does seem to be an important theme initiator for the Mandates, but we do not have enough data to draw conclusions. The methodology could be used to solve this problem.

**Does the first section of Similitude 5 have stronger links with the Mandates than with the Similitudes?**

From the diagram showing the flow of major themes, S5.1 only links with other Similitudes. The answer to this question is a cautious 'No'.

The value of the method of Simple Lexical Repetition being able to inform a wide range of applications is therefore obvious. Having noted what issues Hermas thought important, we next look at those issues that Hermas sees as threats. In order to do this we must look at the conflict language in *The Shepherd.*
Chapter 11
CONFLICT AND CONFLICT LANGUAGE

Introduction

This chapter develops a method that will help to highlight the sections of text in *The Shepherd* that deal with conflict. As with the previous chapters of this thesis, the method is a grammar-based one that could be described as a local grammar of conflict. It is necessary, in the first instance however, to consider the meaning of conflict in a sociological sense. The main ideas that are currently being debated will be given, and a recent conflict correspondence used to illustrate conflict language. The second part of the chapter defines the conflict grammar for the early Greek correspondence of 1&2 Corinthians and then the method is applied to *The Shepherd*, which has never received such an analysis before.

Conflict

Conflict occurs wherever humankind forms itself into social groups. It transcends social and cultural boundaries and even the boundary of time, in the sense that it is probably a trait that has existed since primitive times. With the rise of a science of social behaviour in the last century, conflict was always regarded as a negative influence that destabilised society by aggressive, hostile behaviour. Conflict needed to be resolved in order to maintain social equilibrium. Recently\(^ {139}\), however, these views have been challenged and conflict is considered to be, in some cases at least, a constructive process helping to create the identities of groups. According to this view therefore, conflict is a socialization process. Whereas this statement is true, it would be wrong to assert, or imply, that all conflict is positive in its outcome.

The term ‘conflict’ has associated with it all of the negative connotations with which the last two centuries of the post-modern world has invested it. Therefore, there have been a number of attempts by linguists to re-define the terminology by removing the term completely. Schiffrin\(^ {140}\) prefers to use the word ‘argument’ rather than conflict or disagreement. Adopting this terminology allows us to concentrate on

\(^{139}\) Simmel, G. “Conflict” in *Conflict and the Web of Group Affiliations*. K. Wolff (ed.)


208
the problem from a rhetorical point of view and to consider the persuasiveness of the argument.

Arguments can occur in conversation and in writing; in the study of the early church only the latter remain as the evidence of lively debates that probably occurred in many Christian gatherings. The task therefore becomes the assessment of the argument developed by a single writer who has been involved in the dispute, and who now tries to formalise the argument in written form. It becomes immediately obvious that the rhetorical persuasiveness of the argument is thus 'reduced to' a process of logical reasoning. The term 'reduced to' does not mean to imply that it has become something inferior to the original. Yet oral argumentation tends to be spontaneous and can lead off in all sorts of tangential arguments, whereas the written argument tends to be more consistent and focussed. While it is the written, logically reasoned, argument that is found in the Pauline letters, *The Shepherd* is written in such a way that dialogue is part of the narrative. However, it is clear that the argument is still a logical one, not a true dialogue argument, because it is the product of a single writer.

Arguments usually have a beginning, middle and sometimes an end. Rips further qualifies the term 'argument' to include the fact that it must be based on some substantive content. He therefore distinguishes argument (which has substantive content) from mere name-calling (which has no substantive content). Schiffrin also characterises argument as sustained disagreement with the participants competing for 'negotiable goods'. The 'negotiable goods' can include anything, from the physical elements, such as property, to ideologies such as status, philosophies and ideas. Argument will therefore have at its heart a competition for something. Stein and Albro suggest that most arguments are centred on the goal relating to issues of power, control and dominance, however, the danger with this viewpoint is that of over-generalizing the power issue. As Briggs points out:

"Clearly, a number of problems can emerge from uncritical adoption of these insights... One pitfall revolves around a tendency to identify all cultural forms as instances of power relations and to see power as being everywhere the same; historical and cultural specificities often fall victim to a universalistic and functionalist deus ex machina."

This highlights a very important issue that affects those who apply social-

143 Briggs, C.L. (ed) *Disorderly Discourse*. p.6
scientific methodology to the study of early Christian communities. It is that of the methodology itself. The debate about methodology falls into two camps:

• First there are those who, like Horrell, would urge caution in applying models based on current social-science to the ancient Mediterranean world. It would certainly be legitimate to question the relevance, for example, of Schifffrin’s study of American-Jews of the 20th C if the findings of that study were to be used in the investigation of conflict amongst early Christian groups in the 1st century. While Horrell is not suggesting that present day models should be ignored, he is certainly arguing against the over-generalization that a model necessarily adopts. Models very often assume that human behaviour is both predictable and regular.

It is certainly possible to adopt a model-based approach to the social-scientific interpretation of the New Testament, using a model or models derived from other social contexts or historical periods...However, this approach has serious weaknesses and dangers, both at the theoretical/philosophical level and in its tendency to impose the model upon the evidence; a merely pragmatic assertion that the model ‘works’ cannot obviate these deeper problems.144 (bold mine)

Horrell’s concern, then, is that models can be easily misused and can therefore lead to biased results. If possible, he would like to replace the model with the term ‘theoretical framework’.

• Second, there are those who advocate a model-driven approach. Esler supports this approach and in writing a review of Horrell’s book, *The Social Ethos of the Corinthian Correspondence* chose to deal mainly with the first chapters where Horrell expresses his opinion concerning methodology. Esler145 regards models as part of the human process of understanding and he rejects146 the claim made by Horrel that his view of human culture is that human behaviour is predictable and regular.

Elliott147 goes even further and argues that models are useful in order to dictate what material is to be examined as ‘relevant data’. Both Esler and Elliott regard it as essential to state clearly at the outset of an investigation the nature and function of the model to be used. Both researchers, and others who follow them in this approach

are dependent on Carney\textsuperscript{148} who argued:

\textit{A model...acts as a link between theories and observations. A Model will employ one or more theories to provide a simplified (or an experimental or a generalized or an explanatory) framework which can be brought to bear on some pertinent data.}

Thus the conflict between Horrell and Esler can be seen to be an argument about ‘methodology’. It is conducted in a logical written form, each participant showing the flaws in the argument of the other and formulating their own position. Horrell thinks his methodology of drawing the conclusion from the data is better than Esler’s method of using a model to define which data is important. It is possible to observe which ideas Horrell holds strongly enough to defend, and how he shows that his views are strongly held by the manner in which he writes. For the linguist this is Horrell’s \textbf{stance}.

Having introduced the idea of stance, this chapter develops the linguistic assessment of written texts that indicate they were written to answer a situation of conflict.

\textbf{Linguistic Analysis Of Conflict}

The linguistic mechanisms used by writers\textsuperscript{149} to convey personal feelings and assessments have been given different labels, including ‘evaluation’, ‘intensity’, ‘affect’, ‘hedging’ and ‘stance’. They each provide a facet of the writer’s viewpoint in an argument. If we initially concentrate on the idea of stance, then stance is defined as \textit{the linguistic method that a participant in argumentation uses in order to express their relationship to the propositions expressed by the other participants}.\textsuperscript{150}

The stance of the participant is shown in the construction and use of sentences and grammar. It is possible to identify words that can be described as ‘stance markers’ because they are commonly used in expressions of conflict. Precht\textsuperscript{151} shows that only a few of all the possible stance markers are regularly used by any individual (out of

\textsuperscript{148} Carney, T.F. \textit{The Shape of The Past: Models and Antiquity}. The reference to this publication is taken from Elliott.

\textsuperscript{149} And speakers. Since this study concentrates on written evidence only, all that follows will be confined to the role of writer, but it is important to note that such studies have been used in the \textit{modern} context of conversational argument also.


211
approximately 1400 stance markers, we regularly use only about 150), and that the use of those stance markers is shaped by culture and custom.

Biber\textsuperscript{152} suggested that stance consists of three inter-related elements that are:

- **Epistemic stance.** This indicates the degree of certainty a writer has about the information or proposition used in argumentation.

- **Attitudinal stance.** This indicates how the writer feels, or the judgements expressed, about what is written.

- **Style stance.** This indicates how something is written.

Conrad and Biber\textsuperscript{153} show how grammatical devices can be used assess stance. Although they realised that modal verbs and nominal expressions could be important to express stance, they concentrated on adverbial markers. Using the *Longman Spoken and Written English Corpus* they analysed conversation, academic writing and news reportage in order to discover any patterns of use for the adverbial stance markers. Their results indicated that epistemic stance could be broken down into several sub-classes, which they defined as:

a) Degree of certainty or doubt concerning a proposition.

b) Indicating the reality or actuality of a proposition.

c) Indicating the precision of a proposition.

d) Indicating the source of a proposition.

f) Indicating the limitations of a proposition.

Biber\textsuperscript{154} has carried his research further to include modal and semi-modal verbs, adverbials and complement clauses as three major grammatical systems to express stance. This work also includes patterns of change in social norms and changes over time and from which he is able to develop the major grammatical devices that can be used to express stance. This is a first step in the process of obtaining a grammar of conflict. The major grammatical devices are:

\begin{itemize}
  \item [152] Biber, D., Johansson, S., Leech, G., Conrad, S., and, Finegan, E. *Longman Grammar of Spoken and Written English.*
  
  
\end{itemize}
Modal verbs and semi-modal verbs.

Stance adverbials.

Stance verb + 'that' clause.

Stance verb + 'to' clause.

Using the first couple of paragraphs of Horrell's reply to Esler will illustrate the method. The passage was 'part-of-speech tagged' using CLAWS5\textsuperscript{155} (the tagged passage can be found in Appendix 11.1)

"In a recent issue of The Journal of Theological Studies, Philip Esler paid me the honour of writing a lengthy review (Esler 1998a) of my book, The Social Ethos of the Corinthian Correspondence (Horrell 1996; hereafter SE). In that review, rather than focus on the exegetical sections of the book, Esler chose to concentrate on the theoretical issues' which were raised in Part I of the study, the methodological discussion (Esler 1998a: 255). Esler's decision to focus on these issues of theory and method makes a response of wider significance than might otherwise have been the case. A number of recent writers, reviewing the state of social-scientific interpretation of the New Testament, have highlighted the need for discussion of the methodological and theoretical issues that underlie the exegetical practice. Susan Garrett, for example, insists, 'It is...increasingly urgent that scholars of Christian origins engage in sustained reflection on the philosophical implications of the perspectives and models that they choose to employ' (Garrett 1992: 93). Dale Martin speaks of 'a spectrum of opinion about what precisely a social-scientific method should be', with much of the debate, in his view, centring 'on the use of "models" derived from the social sciences' (Martin 1993: 107). Furthermore, like Garrett, Martin maintains that the debate reflects 'profound theoretical differences, particularly epistemological ones' (Martin 1993: 109-10). Discussion and reflection are particularly important to foster and promote dialogue among those who seek to make use of the social sciences in New Testament interpretation, especially given the sometimes rather firm division between two groups who have come to be referred to as 'the social historians' and 'the social scientists'. In the following essay, I shall attempt to respond to Esler's arguments concerning the use of models, the validity of Berger and Luckmann's theory of the social construction of reality, and, finally and more briefly, the major argument of SE.

1. The Use of Models.

In SE I took issue with a widely represented view that a social-scientific approach to the New Testament essentially involves the use of 'models' (SE, 9-18). I argued - and will not simply repeat those arguments here - that there were a number of problems with such a conception of social-scientific method and maintained that it was certainly not the only way to approach the New Testament 'sociologically'. Not the least of these problems, it seems to me, is the way in which a model-based approach can lead to historically..."
and culturally variable evidence being interpreted through the lens of a
generalized model of social behaviour. The rich diversity of human
behaviour is thereby homogenized and explained in terms of what is 'typical'
which can actually be no real explanation at all (cf. SE, 287-89). Although
Esler is conscious of these dangers, and insists on the need for sensitivity to
cultural variation and specificity (see Esler 1998b: 4-5), his most recent
work still exhibits the problem, as will be shown below.

In the above text, modal verbs are highlighted in yellow; adverbs are highlighted in
blue; 'to' markers highlighted in green; 'that' markers highlighted in purple.

Horrell begins with a polite introduction, “Philip Esler paid me the honour of
writing a lengthy review (Esler 1998a) of my book, The Social Ethos of the
Corinthian Correspondence” designed to show his respect of his opponent and to
remind the reader that the review was of Horrell’s book. The conflict is initiated by
the fact that Esler devotes most of the review, not in evaluation of the book, but in
criticism of Horrell’s call for a rethink of methodological approaches to New
Testament study, “Esler’s decision to focus on these issues of theory and method
makes a response of wider significance than might otherwise have been the case.”
Horrell then defends his position by reminding the reader that he is not the only one
to hold these views, “A number of recent writers, reviewing the state of social-
scientific interpretation of the New Testament, have highlighted the need for
discussion.” He is deliberately vague in his comment that ‘a number of recent
writers’ can be called on for support. Horrell therefore believes he has strong
backing for his case, but he mentions two in particular, Susan Garrett who also thinks
it ‘increasingly urgent’ that discussion should take place, and, Dale Martin who also
urges debate because the question of models is “particularly epistemological”.
Having declared his position is not an isolated one, Horrell declares, “I shall attempt
to respond to Esler’s arguments”.

Horrell’s stance on the question of social-scientific models is clearly that of
certainty, “I took issue with a widely represented view” and “I argued.. that there
were a number of problems”. He sees himself as a champion of an alternative view
to Esler, who, Horrell implies sticks rigidly to the position that models are the only
way to study the New Testament sociologically. Horrell proposes the counter
argument, “it was certainly not the only way to approach the New Testament
'sociologically'. Model based methods have the danger of producing results "which can actually be no real explanation at all". This in itself is a damning assertion about Esler's methodology, but Horrell then goes for a knock-out by suggesting to the reader that Esler knows there are problems with the method he adopts, "Although Esler is conscious of these dangers", but, "his most recent work still exhibits the problem".

Although this is a short and densely substantive passage, it demonstrates how the stance markers suggested by Biber can highlight the important sentences in a written, logical argument.

This method is now adapted in order to identify major lines of argument in GREEK texts of the early 1st century. This is exactly the kind of matter that Horrell argues against! Will stance markers that give insight into 20th century British and American culture be valid markers in a 1st century Greek culture? The next section will assess the possible use of stance markers in the conflict letters of 1 and 2 Corinthians. It will then be possible to compare results obtained by this method with the results obtained by Horrell, as reported in his book, The Social Ethos of the Corinthian Correspondence.

Conflict in Paul's Corinthian Correspondence

The authorship of 1 Corinthians is not disputed as being essentially Pauline. Some minor phrases and comments are considered to be interpolations, otherwise authorship is undisputed. Similarly, the authorship of 2 Corinthians is not disputed, apart from the section 2 Corinthians 6:14-7:1 which has sometimes been doubted as to its Pauline origin. Most of the contention lies in the determination of the number of letters that Paul wrote to the community, and whether these letters can be reconstructed from the two letters that we now have. Details of these arguments can be found in any commentary on 1&2 Corinthians, but discussion of this is not the purpose of the following analysis. The letters, however many there may be, are Paul’s response to the conflict that was in progress at Corinth. Can we, then, use Biber's local grammar of stance to determine Paul’s position with respect to the other participants in the conflict? Adjustments will need to be made to Biber’s lexical categories since they cannot be carried over from English to Greek in an easy

fashion. We are particularly concerned with modal and semi-modal verbs, stance adverbials, ‘that’ complement clauses and ‘to’ complement clauses.

Greek does not have modal verbs in the way that English does; rather Greek verb mood expresses certainty etc. When a Greek sentence expresses a condition, such as a wish, an obligation etc. these conditions are sometimes referred to as ‘modalities’. Modalities are expressed by a variety of means, for example, conditional sentences can be introduced with εἰ or ἔδιν. Furthermore, conditional sentences can be sub-divided into ‘factive’ and ‘non-factive’ (keeping the nomenclature of Biber). Obligation is expressed by the impersonal verb δεῖ (sometimes ἐξεστὶ).

Adverbs can be found in a text quite easily. Not all adverbs will function as stance adverbials; therefore, each adverb will need to be evaluated in its context.

Complementary ‘that’ clauses will be introduced by ὅτι.

Infinitive clauses, ‘to’, can also be found easily, but like adverbs, each one will need to be evaluated in context.

Using the parsed New Testament text of 1Corinthians the grammatical markers mentioned above are found. Those markers that relate to authorial stance were tabulated and assessed. In the table below, verses are grouped together and colour-coded into similar themes. The column labelled ‘Conflict’ is a summary term for the theme of the verse in that group.

<table>
<thead>
<tr>
<th>REF</th>
<th>STATEMENT</th>
<th>CONFLICT</th>
</tr>
</thead>
<tbody>
<tr>
<td>01:11</td>
<td>ἐδηλώθη γάρ μοι --- ὅτι ἔριδες εἰν ὑμῖν εἰσιν</td>
<td>Conflict initiation: Divisions</td>
</tr>
<tr>
<td>01:14</td>
<td>εὐχαριστῶ (τῷ θεῷ) ὅτι οὐδένα ὑμῶν ἐβάπτισα</td>
<td>Baptismal Groups</td>
</tr>
<tr>
<td>01:15</td>
<td>ἵνα μὴ τις εἶπῃ ὅτι εἰς τὸ ἐμὸν ὄνομα ἐβαπτίσθητε</td>
<td></td>
</tr>
<tr>
<td>01:17</td>
<td>ὅ γαρ ἀπέστειλεν με Ἑρωτός βαπτίζειν ἅλλα εὐαγγελίζεσθαι</td>
<td></td>
</tr>
<tr>
<td>01:21</td>
<td>ὁ θεὸς διὰ τῆς μορίας τοῦ κηρύγματος σώσαι</td>
<td>Wisdom &amp; Boasting</td>
</tr>
<tr>
<td>01:26</td>
<td>βλέπετε γάρ --- ὅτι οὐ πολλοὶ σοφοί</td>
<td></td>
</tr>
</tbody>
</table>

157 Obtained from www.byztxt.com, Westcott-Hort Parsed Text with NA27 Variants, WHN27PRS: The Westcott-Hort edition of 1881 with complete parsing information for all Greek words. Readings of Nestle27/UBS4 shown as footnotes, also with complete parsing information attached (footnote format uses the piping symbol). In this text ALL accents and breathings are removed.
<table>
<thead>
<tr>
<th>Time</th>
<th>Greek Text</th>
<th>English Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>01:29</td>
<td>ὀπως μὴ καυχήσηται πᾶσα σάρξ εἰνώπιον τοῦ θεοῦ</td>
<td>Declare that there is no adulterer among you.</td>
</tr>
<tr>
<td>02:02</td>
<td>οὐ γὰρ ἐκματά τι εἰδέναι ἐν ὑμῖν εἰ μὴ Ἰησοῦν Χριστὸν</td>
<td>You have no need to be instructed about love, for Christ is meant.</td>
</tr>
<tr>
<td>03:02</td>
<td>γάλα ὑμᾶς ἐπότισσα ὦ βρώμα ὑπὲρ γὰρ ἐδύνασθε ἀλλ’ ὦδε ἔτι νῦν ὀδύνασθε</td>
<td>You are feeding on milk, for you are now able to eat meat.</td>
</tr>
<tr>
<td>03:03</td>
<td>ἔτι γὰρ σαρκικὸν ἐστε ὅπου γὰρ ἐν ὑμῖν ζήλος καὶ ἐρίς</td>
<td>For you are now living like flesh-eaters when you are living like beasts.</td>
</tr>
<tr>
<td>03:16</td>
<td>οὐκ οίδατε ὅτι ναὸς θεοῦ ἐστε</td>
<td>Did you not know that a temple of God is among you?</td>
</tr>
<tr>
<td>03:20</td>
<td>καὶ πάλιν κύριος γινώσκει τοὺς διαλογισμοὺς τῶν σοφῶν ὃτι εἰσίν ματαιοὶ</td>
<td>And again, the Master knows your thoughts and that they are wicked.</td>
</tr>
<tr>
<td>04:15</td>
<td>εάν γὰρ μιρίους παιδαγωγοὺς ἔχητε ἐν Χριστῷ ἀλλ’ ὦ πολλοῖς πατέρας ἐν γὰρ Χριστῷ Ἰησοῦ διὰ τοῦ εὐαγγελίου ἐγὼ ὑμᾶς ἐγέννησα</td>
<td>If you had a thousand masters in Christ, then you would be sons of many fathers.</td>
</tr>
<tr>
<td>04:19</td>
<td>ἔλευσομαι δὲ ταχέως πρὸς ὑμᾶς εάν ὁ κύριος βλέπῃ καὶ γνώσομαι ὦ τῶν λόγων τῶν πεφυσιωμένων ἀλλὰ τὴν δύναμιν</td>
<td>I shall come quickly to you, if the Master sees and knows you and your works.</td>
</tr>
<tr>
<td>05:01</td>
<td>ὅλως αἰκουέται ἐν ὑμῖν πορνεία --- ὅπως γυναῖκα τινα τοῦ πατρὸς ἐξεύρην</td>
<td>Always let yourselves be guilty of淫欲 --- as though a woman of the Father has found something.</td>
</tr>
<tr>
<td>05:03</td>
<td>ἐγὼ --- ἡδη κέκρικα ὡς παρών τόν οὕτως τοῦτο κατεργασάμενον</td>
<td>I have already determined what I am doing to the Galatians.</td>
</tr>
<tr>
<td>05:06</td>
<td>οὐκ οίδατε ὅτι μικρά ζήμη ὅλον τὸ φύραμα ζύμοι</td>
<td>Did you not know that a small stain makes the whole dough bad?</td>
</tr>
<tr>
<td>05:09</td>
<td>ἐγραφαία ὑμῖν ἐν τῇ ἐπιστολῇ μὴ συναναμίγνυσθαι πόρνους</td>
<td>The scriptures have written that you should not associate with prostitutes.</td>
</tr>
<tr>
<td>05:11</td>
<td>ὑπὸν ὒπερ ἐγραφαία ὑμῖν μὴ συναναμίγνυσθαι ἐὰν τίς ἀδελφὸς ὑμῶν ὁμοφρόμοιος --- τῷ τοιούτῳ μηδὲ συνεδθῆτε</td>
<td>If the scriptures have written that you should not associate with prostitutes, if one of your brethren is a fornicator --- to the same one, do not even eat with him.</td>
</tr>
<tr>
<td>05:12</td>
<td>τοῖς γὰρ μου τοὺς ἐξω κρίνειν</td>
<td>For those are mine, and I will judge them.</td>
</tr>
<tr>
<td>06:01</td>
<td>τολμᾶ τις ὑμῶν πράγμα ἐξων πρὸς τὸν ἐτερον κρίνεσθαι ἐπί τῶν ἁδικῶν</td>
<td>Do not let any of you judge one another.</td>
</tr>
<tr>
<td>06:02</td>
<td>ἢ οὐκ οίδατε ὅτι οἱ ἁγιοὶ τῶν κόσμων κρυνόθησιν</td>
<td>Did you not know that the righteous of the world hide themselves?</td>
</tr>
<tr>
<td>06:03</td>
<td>οὐκ οίδατε ὅτι ἁγγέλους κρυνόμενον</td>
<td>Did you not know that angels hide themselves?</td>
</tr>
<tr>
<td>06:04</td>
<td>βιωτικά μὲν οὖν κριτήρια ἔχετε ἐξήτευτε τοὺς εξουδετερούμενους ἐν τῇ ἐκκλησίᾳ τούτους καθέστε</td>
<td>You have been wise judges, to have rejected the defiled ones in this church.</td>
</tr>
</tbody>
</table>
Immorality And The Body

Marriage

Idols

Apostleship
<table>
<thead>
<tr>
<th>Time</th>
<th>Greek Text</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:04</td>
<td>μὴ οὐκ ἔχομεν ἐξουσίαν φαγεῖν καὶ πείνα</td>
<td>Idolatry</td>
</tr>
<tr>
<td>09:05</td>
<td>μὴ οὐκ ἔχομεν ἐξουσίαν ἀδελφῆν γυναῖκα περιέχειν</td>
<td>Charismata</td>
</tr>
<tr>
<td>09:06</td>
<td>ἦ μονὸς ἐγώ καὶ βαρναβᾶς οὐκ ἔχομεν ἐξουσίαν μὴ ἐργαζομέναι</td>
<td></td>
</tr>
<tr>
<td>09:13</td>
<td>οὐκ ὀδηγάτη ὅτι οἱ τὰ Ιερὰ ἐργαζόμενοι τὰ ἐκ τοῦ Ἱεροῦ ἐσθίουσιν</td>
<td></td>
</tr>
<tr>
<td>09:14</td>
<td>τοῖς τῷ εὐαγγέλιον καταγγέλλουσιν ἐκ τοῦ εὐαγγελίου ἦν</td>
<td></td>
</tr>
<tr>
<td>09:15</td>
<td>γὰρ μοι μάλλον ἀποθανεῖν ἢ τὸ καύχημα μου οὐδεὶς κεινῶσί</td>
<td></td>
</tr>
<tr>
<td>09:16</td>
<td>εάν γὰρ εὐαγγελίζωμαι οὐκ ἐστίν μοι καύχημα ἀνάγκη γὰρ μοι ἐπίκειται οὐαί γὰρ μοι ἐστίν εάν μὴ εὐαγγελίζωμαι</td>
<td></td>
</tr>
<tr>
<td>09:24</td>
<td>οὐκ ὀδηγάτη ὅτι οἱ ἐν σταδίῳ τρέχουσε πάντες μὲν τρέχουσιν εἰς δὲ λαμβάνει τὸ βραβείον</td>
<td></td>
</tr>
<tr>
<td>10:01</td>
<td>εὐγό τούνων οὕτως τρέχω ως οὐκ ἄδηλως</td>
<td></td>
</tr>
<tr>
<td>10:01</td>
<td>οὗ θέλω γὰρ ύμᾶς ἀγνοεῖν --- ὅτι οἱ πατέρες ἡμῶν πάντες ὑπὸ τὴν νεφέλην ἦσαν καὶ πάντες διὰ τῆς θαλάσσης δηλθοῦν</td>
<td>Idolatry</td>
</tr>
<tr>
<td>10:21</td>
<td>οὗ δύνασθε ποτήριον κυρίου πίνειν καὶ ποτήριον δαιμονίων οὐ δύνασθε τραπεζίης κυρίου μετέχειν καὶ τραπεζίης δαιμονίων</td>
<td></td>
</tr>
<tr>
<td>10:28</td>
<td>εάν δὲ τις ύμῖν εἰπη τοῦτο ἱεροῦ ὃν ἐστίν μὴ ἐσθίετε</td>
<td></td>
</tr>
<tr>
<td>11:22</td>
<td>μὴ γὰρ οἰκίας οὐκ ἔχετε εἰς τὸ ἐσθίειν καὶ πίνειν --- τι εἰπὼ ὑμῖν ἐπαινεῖσθαι ύμᾶς ἐν τούτῳ οὐκ ἐπαινῶ</td>
<td></td>
</tr>
<tr>
<td>12:01</td>
<td>περὶ δὲ τῶν πνευματικῶν ἀδελφοὶ οὗ θελὼ ύμᾶς ἀγνοεῖν</td>
<td>Charismata</td>
</tr>
<tr>
<td>14:05</td>
<td>θέλω δὲ πάντας ύμᾶς λαλεῖν γλώσσας μᾶλλον δὲ Ἰνα προφητεύοντες</td>
<td></td>
</tr>
<tr>
<td>14:06</td>
<td>ἰν νῦν ἄδελφοι εάν ἐλθὼ πρὸς ύμᾶς γλωσσαῖς λαλῶν τί ύμᾶς ωφελῆσθω</td>
<td></td>
</tr>
<tr>
<td>14:08</td>
<td>καὶ γὰρ εὰν ἄδηλον σάλπιγξ φωνῆν δοῦ τίς παρασκευάσεται εἰς</td>
<td></td>
</tr>
</tbody>
</table>
From this table we can see that Paul deals with these major issues at Corinth:

- Divisions based on baptismal groupings.
- Those who boasted about their wisdom.
- Problems about immorality in the church.
- Open conflict displayed in the public law courts.

| 14:23 | εάν οὖν συνέλθη ἡ ἐκκλησία ὅλη ἐπὶ τὸ αὐτὸ καὶ πάντες λαλῶσιν γλώσσαις |
| 14:28 | εάν δὲ μὴ ἢ διερμηνευτὴς σιγάτω |
| 14:30 | ὁ πρῶτος σιγάτω |
| 14:31 | δύνασθε γὰρ καθ' ἕνα πάντες προφητεύειν |
| 14:37 | ἐπιγινωσκέτω δ' γράφω ὑμῖν ὅτι κυρίου ἐστὶν ἐντολή |
| 15:03 | παρέδωκα γὰρ ὑμῖν --- ὅτι Χριστὸς ἀπήγαγεν ὑπὲρ τῶν ἀμαρτιῶν καὶ ὅτι ἐτάφη καὶ ὅτι ἐγήγερται τῇ ἡμέρᾳ τῇ τρίτῃ κατὰ τὰς γραφὰς |
| 15:05 | καὶ ὅτι ὄφης |
| 15:08 | εἰ δὲ Χριστὸς κηρύσσεται ὅτι ἐκ νεκρῶν ἐγήγερται πῶς λέγουσιν ἐν ὑμῖν τινες ὅτι ἀνάστασις νεκρῶν οὐκ ἐστιν |
| 15:25 | δεὶ γὰρ αὐτῶν βασιλεύειν τούτῳ δὲ φημὶ ἀδελφοὶ ὅτι σάρξ καὶ αἷμα βασιλείαν θεοῦ κληρονομησάτως οὐ δύναται οὐδὲ ἡ φθορά της ἀφισχοῦται κληρονομεῖ |
| 15:50 | δεὶ γὰρ τὸ φθαρτὸν τοῦτο ἐνύστασθαι ἀφθαρσίαν |
| 15:53 | εἰδότες ὅτι ὁ κόπος ὑμῶν οὐκ ἐστὶν κεῖσθαι ἐν κυρίῳ |
| 15:58 | οὖ θέλω γὰρ ὑμᾶς ἀρτί ἐν παρόδῳ ἰδεῖν ἐλπίζω γὰρ χρόνον τινὰ ἐπιμεῖναι πρὸς ὑμᾶς εάν ὁ κύριος ἐπιτρέψῃ |
| 16:07 | Paul's Preaching |
| 15:12 | Paul's Wish to Visit |
| 15:50 | Death and Resurrection |

**Paul's Preaching**

**Death and Resurrection**

**Paul's Wish to Visit**
• The role of marriage.
• Idolatry and freedom.
• Questions about Paul’s apostleship.
• Correct use of charismata in the church.
• The substance of Paul’s preaching and the resurrection.
• Paul’s wish to visit the Corinthian church.

Dealing with these findings very briefly, since this thesis is not about the Corinthian situation, it could be argued that the problems in Corinth were widespread and ingrained.

The factions that were formed around certain named leaders in 1 Cor 1:12 seems to be in Paul’s mind when the problems of division had been reported. Commentators have given various suggestions about the nature of these divisions, ranging from scepticism about their existence to the struggle between patrons in the church. While it is more likely that some catalyst for division was present, it would appear from the table above, that this division is related in some way to baptism and those who had been baptized. These parties could therefore have been formed around itinerant teachers who taught and baptised at Corinth. Horrell mentions baptismal rites as a source of conflict, but not in relation to itinerant teachers.

There were in Corinth those who considered themselves to be wise and who probably formed an elite group. These probably disparaged Paul and his teaching, for later we find him defending both these aspects of his calling. Chow mentions the work of Horsley and his suggestion of their interest in Hellenistic-Jewish traditions related to devotions to heavenly Sophia. Horrell takes a more cautious view, “in my view it is most likely to have come from their own socio-religious background”, but he also thinks that they despised Paul’s lack of rhetorical skills. It is interesting to see the number of times Paul uses the expression, oũk olı̂sατε ὅτι... as highlighted in the tabular analysis above. This must surely

159 Theissen, G. The Social Setting Of Pauline Christianity. quoted in Chow, J.K. Patronage and Power. p.94
160 Horrell, D.G. Social Ethos. p.82
161 Chow, K.K. Patronage and Power. p.118
162 Horrell, D.G. Social Ethos. p.121
not be coincidental but sarcasm!

Fiorenza and Conzelmann\textsuperscript{163} classify the next three problems together, that is, the problem of litigation is probably tied up with marriage law, even to the case of the relationship of son and stepmother. Horrell\textsuperscript{164} also mentions the question of litigation. Paul’s stance is clearly displayed as being opposed to such actions, and he even encourages the Corinthian Christians to follow his example (remaining celebate).

Idols and idolatry are seen to form two sections in the above table, the first from a philosophical point of view, and the second from the practical point of food offered to idols. Horrell\textsuperscript{165} believes this to be a class-based struggle.

Paul strongly defends his apostleship, and it would appear that there were those in the Corinthian church who must therefore have launched a personal attack against Paul. Horrell also sees this as a major theme, devoting a chapter to it in his book\textsuperscript{166}. There is no need to repeat the arguments here, but again the analysis of 1 Corinthians for conflict language has highlighted this issue.

Horrell notes that there have been those who have maintained that the exercising of certain charismatic gifts is related to status, but he hedges on this point, “But even if we do not assume that the tongue-speakers were drawn from any particular social group…”\textsuperscript{167}. Chow, on the other hand, notes those who have suggested the influence of Hellenistic enthusiasm, but he also hedges by saying, “Such a characterization of the Corinthian outlook may need to be refined at different points”\textsuperscript{168}. Both Horrell and Chow comment on the potentially divisive nature of the group, even if they cannot agree about its background.

Horrell deals with the section on Paul’s preaching in a section of his book entitled “The theology of Paul’s Corinthian ‘opponents’”\textsuperscript{169}. Clearly Paul felt that he needed to remind the Corinthians of the message that he had delivered to them, which was in danger of being lost in the clamour at Corinth.

Finally Paul expresses his wish to visit them, perhaps because there were those who had remarked on his long absence.

\textsuperscript{163} quoted in Chow, K.K. \textit{Patronage and Power}, p.132
\textsuperscript{164} Horrell, D.G. \textit{Social Ethos}, p.109
\textsuperscript{165} Horrell, D.G. \textit{Social Ethos}, p.107
\textsuperscript{166} Horrell, D.G. \textit{Social Ethos}, chapter 5
\textsuperscript{167} Horrell, D.G. \textit{Social Ethos}, p.177
\textsuperscript{168} Chow, K.K. \textit{Patronage and Power}, p.117
\textsuperscript{169} Horrell, D.G. \textit{Social Ethos}, p.119
It can be observed therefore, that analysis of the text for conflict language, based on Biber’s theory, seems to guide us to all of the causes for conflict that Horrell finds in his work. The transposition of the model through time and language appears to work well.

**Conclusion**

A local grammar of conflict for Greek texts has been proposed in this chapter and has been found to offer the advantage of focussing attention on the relevant issues in that conflict. It is not meant to be a thorough examination of Greek conflict language, but a tool to guide to relevant issues in a large text, such as *The Shepherd*. Ideally, the analysis should be carried out on a large corpus of Greek texts, as Biber did on English texts. **This highlights the urgent need for an agreed corpus of Greek texts that could be used in linguistic studies. Such a corpus would offer the advantage of a fixed set of data by which to judge any new linguistic or statistical theories.**

Having shown that the method of identifying conflict in texts works, in principle, it will now be used and refined in the next chapter in order to discover the concerns of Hermas in *The Shepherd*. 
Chapter 12

CONFLICT LANGUAGE IN THE SHEPHERD OF HERMAS

Introduction

The previous chapter suggested that it was possible, in principle, to discover conflict language in a Greek text. This was based on the fact that Biber’s theory of a grammar of conflict could be transposed through language and time. It makes no sociological judgement as to the nature of conflict, only that some elements of grammatical usage in conflict discourse may be similar. The texts used in that analysis were, a) a Western English language academic text and b) a Mediterranean 1st century Greek epistle. These were similar in the mode of argument used, that is, both developed logical strings of evidential material that culminated in the proof of their argument. The sociological differences in the nature and manner of the arguments across both a cultural and a temporal divide would certainly need to be discussed along with any linguistic theory. This was not done in the previous chapter however, because 1 Corinthians is not the topic of this thesis. In the present chapter there will be need to conduct a more thorough-going analysis that includes a sociological assessment of 1st century conflict.

When we come to the study of The Shepherd, there is also the problem of genre difference compared to the texts studied in the previous chapter. It has already been noted (in the first chapter of this thesis) that The Shepherd cannot be considered as a text of a well planned logical argument. Rather, it is known for its rambling (some might say, boring) content and style. It is not apocalyptic, although it contains a visionary guide; it is not an epistle, although parts are to be written and circulated to the churches; it is not a sermon, although it encourages and exhorts and offers advice on Christian conduct; it is not philosophically theological, or theologically philosophical, although it offers some confessional formulae. Perhaps, since it does not seem to fit into any particular genre, it is best described as ‘a Christian romance story’. One thing is obvious however, it has a narrative structure. Therefore, the analysis of the text for conflict language will require a greater degree of scrutiny with regard to the function of the conflict within the story. The analysis of the text in this chapter we make use of two approaches, Discourse Analysis and Conflict Language.

This chapter is therefore a mixture of a local grammar of conflict, discourse analysis (particularly narrative semiotics and distinction theory analysis) and it
concludes with discussion of some themes from a socio-rhetorical viewpoint.

**Narrative Semiotics**

Narrative semiotics is a method of analysis that seeks to characterise the narrative on two levels.

First the text has a 'surface structure'. This is a concern with the structure of the words and syntax used in the narrative.

Second the text has a 'deep structure'. The text will have a system of values, norms and attitudes embedded in the narrative. The deep structure is obviously related to the social values of the writer and reader.

In order to identify the underlying values, narrative semiotics uses the model of a semiotic square. A square is divided into four segments, as shown in Fig. 12.1.

![Fig. 12.1](image)

*Fig. 12.1 Typical construction of a semiotic square. The symbols are defined in the text.*

The square is a visual means by which a 'compare and contrast' of topics in the text can be logically set out. The relationship between S1 and S2, and S1 and S3, is that of opposites. The relationship between S1 and S4, and S2 and S3, is one of contradiction.

This is a common method used in problem solving logic. The problem is sharply focussed by defining what the problem is, and what the problem is not. In
the symbolic representation of the square, S1 would be what the problem is, S4 would be what the problem is not. In narrative S1 would be a theme, S4 would be what the theme is not.

Since the concern of this chapter is that of 'conflict' expressed in the text of *The Shepherd*, it is possible to construct a semiotic square that represents the issue of participants in conflict. A reading of the text shows that there is a conflict between the Christian and the pagan world, its moralities and its dangers. We can therefore say that S1 = The pagan community (the world). S4 is therefore the contradiction of a world of paganism, that is, S4 = The righteous community (the church).

S2 is the opposite of S1. If S1 is the physical and immoral, represented by the community of the world, then S2 = the spiritual and ideal world, represented by spiritual figures, that is, S2 = The Visionary Figures. These Visionary Figures include the 'heavenly Rhoda', 'The Lady Church', 'The Shepherd' and 'The Seven Maidens'.

S3 is the contradiction of S2, that is, S3 = Hermas.

This is expressed in the form of a semiotic square in Fig. 12.2.

![Fig. 12.2 Typical construction of a semiotic square. The symbols are defined in the text.](image)

The text can therefore be analysed for conflict in these four areas of concern. Our primary interest is to discover how Hermas (the assumed author) interacts with these four levels. The data found in the text will be biased to Hermas' point of view, but any conflict that has occurred will be discernable in the text. Just as when a stone is thrown into water, it leaves evidence of its encounter with the water by causing ripples on the surface, so any occurrence of conflict that Hermas, or his community, has had with 'The World' or 'The Church' will leave tell-tale signs in
the text. These tell-tale signs are to be found in the local grammar of conflict that
Hermas uses in his conflict with pagan society. Hermas is sometimes seen in conflict
with the Visionary Figures too. He has an internal conflict which causes self-doubt
and despair, as noted for example in V1.2.1. Most importantly for this study, the text
is analysed for conflict language in Hermas’ dealings with the church. Therefore,
attention will be focussed only on that one arena of conflict.

Viewing the diagonals of the Semiotic Square (Fig. 12.2), ‘The Church’ can be
viewed as an idealised projection of The World into a spiritual dimension. Similarly,
‘The Visionary Figures’ are an idealised spiritual projection of the Hermas ego.
Consequently, we see that the narrative involves both psychoanalytic and psycho-
dramatic concepts, and this is in keeping with the work of Greimas170, a leading
figure in the Narrative Semiotic field of study.

Stance and Evaluation

In the previous chapter stance was defined as, *the linguistic method that a participant
in argumentation uses in order to express their relationship to the propositions expressed
by the other participants*. It was also noted that stance can be divided into three categories,
epistemic, attitudinal and style.

Epistemic stance is the comment of the writer on the status of the information in the
main clause. The writer indicates the degree of probability, reality and precision of the
proposition and may also express a comment upon the source and perspective of the
proposition. For example:

\[\text{allá πάντως ἐπὶ τὴν καρδιὰν σοῦ ἀνέβη περὶ αὐτῆς} \quad (V1.2.4)\]

But (the thought) about her *certainly* did enter your heart.

In this example the Visionary Figure indicates the reality of the of the proposition by
using the adverb πάντως - ‘certainly’. Biber terms this a factive adverbial.

Attitudinal stance is sometimes referred to as ‘evaluation’. ‘Evaluation’ and
‘Attitudinal Stance’ are essentially the same thing and can be defined as the writer’s
attitude, viewpoint or feelings about what is written. Evaluation is the expression of the
values and opinions of the writer, which, in turn, may reveal the ideology of the society
that has produced the text. Therefore, evaluation is concerned with the writer’s expression
of attitudes, feelings, judgements, or commitment to what is written. For example:

λέγει γάρ σοι δικαίως ὁ κύριος τῆς χώρας ταύτης ἢ τοῖς
νόμοις μου χρῶ ἢ ἐκχώρει ἐκ τῆς χώρας μου (S1.1.4)

For the lord of this country rightly says to you, ‘Either use my laws or leave
my country.’

In this example the writer evaluates the proposition, ‘Either use my laws or
leave my country’, and judges this to be a fair, right or just proposition.

Style stance is the comment of the writer about the way in which the
proposition is meant to be understood.

Biber defines ‘stance markers’ as aspects of grammar commonly used to
express evaluation and we have seen that these include modal verbs, stance
adverbials, ‘that’ complement clauses and ‘to’ complement clauses. It is important
to remember, and can be seen in the examples given above, that stance meanings are
carried by complete propositions rather than single words. Although it is customary
to refer to stance adverbials, it is the proposition or proposal that is the unit of
analysis.

If the assertion that The Shepherd does not display much logical argumentation
is true, then we would expect to find a greater number of attitudinal stance than
epistemic stance propositions. Consequently, the first stage in the analysis will be to
discover instances of attitudinal stance.

In order to perform this analysis attitudinal adverbs, attitudinal verbs,
attitudinal adjectives and attitudinal nouns will need to be found in the text.

Hunston and Sinclair\textsuperscript{171} utilize Biber’s major grammatical devices used to express
stance, but they refine the grammatical constructions to allow for a programmable,
unsupervised, computer controlled extraction of evaluative statements from a text.
Biber had found that ‘that’ and ‘to’ complement clauses controlled by adjectives or
nouns could act as stance markers. The problem that Hunston and Sinclair needed to
resolve was how to detect ‘stance adjectives’ and ‘stance nouns’. Adjectives and
nouns are open class words that are being added to with time, consequently a list of
attitudinal adjectives and nouns cannot be constructed. Each occurrence of an
adjective or noun has to be examined in its context within the text being studied. If
the researcher has to do the work on his/her own, then it would take a considerable
time to analyse a long text, and would be impossible to perform on a corpus of texts

& Thompson, G. pp. 74-101.
containing many millions of words. Given the correct programming, a computer could do the task very quickly.

Corel Paradox 8 was programmed to perform a similar function on *The Shepherd*, but required a supervised examination of the data following the program run. Because the different stance markers overlap in some instances, that is, a number of stance markers can occur in any single sentence, it is possible to reduce the amount of data handling by considering the sentence once only. The sentences were catalogued according to topic in order that manageable and focussed assessment could be made.

The concern which is of most interest in this chapter is that of Hermas and the church. Hermas expresses a number of evaluative statements about the church, which give some indication of Hermas’ conflict with his fellow Christians in Rome. We can look at these concerns under a number of headings.

**Church Leadership.**

άκουε νῦν περὶ τῶν λίθων τῶν ὑπαγόντων εἰς τὴν οἰκοδομὴν. οἱ μὲν οὖν λίθοι οἱ τετράγωνοι καὶ λευκοὶ καὶ συμφωνοῦντες ταῖς ἁρμογαῖς αὐτῶν οὕτω εἰσάν οἱ ἀποστόλοι καὶ ἐπίσκοποι καὶ διδάσκαλοι καὶ διάκονοι οἱ πορευθέντες κατὰ τὴν σεμνότητα τοῦ θεοῦ καὶ ἐπισκοπήσαντες καὶ διδάχαντες καὶ διακονήσαντες ἁγιῶς καὶ σεμνῶς τοῖς ἐκλεκτοῖς τοῦ θεοῦ οἱ μὲν κεκοιμημένοι οἱ δὲ ἔτι οὖντες. Καὶ πάντοτε ἑαυτοῖς συνεφώνησαν καὶ ἐν ἑαυτοῖς εἰρήνην ἔσχον καὶ ἀλλήλων ἡκούσαν. διὰ τούτο ἐν τῇ οἰκοδομῇ τοῦ πύργου συμφωνοῦσιν οἱ ἁρμογαί αὐτῶν. (V3.5.1)

These words are spoken by the Visionary Figure, The Church, in the narrative in explanation of the stones used in the building of the Tower. The author of the text (Hermas) is using the vision figure as speaker of the words in order that they will appear to have divine authority. These leaders are equated with square, white stones, by which symbol Hermas implies their perfection in fitting into the tower. He evaluates the work of the Apostles, bishops, teachers and deacons very highly, as being ‘according to the majesty of God’. Similarly he evaluates their service to the
church very highly, serving in holiness and reverence, they always agreed, were at peace with one another and listened to one another. Perhaps Hermas is giving an idealised view of the church leaders, or he is looking back to a previous golden age of the church. He comments further about these notable characters that some have fallen asleep (euphemism for death), which implies that some are still alive. Therefore, not all the church leaders at the time of writing still exhibit these qualities that Hermas regards so highly. But it is the closing clauses that seem to be personal to Hermas:

\[
\text{Καὶ πάντοτε ἐαυτοῖς συνεφώνησαν}
\]
\[
\text{καὶ ἐν ἐαυτοῖς εἰρήνην ἔσχον}
\]
\[
\text{καὶ ἄλληλων ἥκουσον.}
\]

Hermas is looking back to the past when things were better than they were currently with Hermas. Perhaps he is talking in this latter instance about the experience he has had, either in his own community, or in the wider context of the Christian communities in Rome. He has not been given a hearing by some in the church. There are arguments and disagreements which Hermas finds to be distressing, and the church is changing to a more worldly and hostile place.

He takes up this theme again in Similitude 9.25.2 where we find the explanation that is given of the eighth mountain. There is a revealing comment:

\[
\text{ἀπόστολοι καὶ διδάσκαλοι οἱ κηρύξαντες εἰς ὅλον τὸν κόσμον καὶ οἱ διδάσκαλοι σεμινῶς καὶ ἁγνῶς τὸν λόγον τοῦ κυρίου καὶ μηδὲν ὁλῶς νοσφισάμενοι εἰς ἐπιθυμίαν ποιηράν ἀλλὰ πάντοτε ἐν δικαιοσύνῃ καὶ ἀληθείᾳ πορευθέντες καθὼς καὶ παρέλαβον τὸ πνεῦμα τὸ ἁγιον. τῶν τοιούτων οὖν ἢ πάροδος μετὰ τῶν ἁγγέλων ἐστίν (S9.25.2)}

Hermas evaluates the preaching and teaching of the Apostles and teachers as being proclaimed reverently (σεμινῶς) and purely (ἀγνῶς). He then remarks of these that they

\[
\text{μηδὲν ὁλῶς νοσφισάμενοi eἰς ἐπιθυμίαν ποιηράν.}
\]

This implies that there were some leaders present in the church who were not only argumentative but were also guilty (at least in Hermas' mind) of νοσφίζεσθαι, the
crime of Ananias and Sapphira (Acts 5:2).

Hermas also commends the ἐπίσκοποι and φιλοξενοι in S9.27.2 who sheltered the destitute and the widows. This contrasts with those mentioned in the paragraph above who were concerned with their own wealth. Again in S9.26.2, Hermas mentions the leaders who misuse their trust, and he clearly announces their fate:

οἱ μὲν τοὺς σπίλους ἔχοντες διάκονοι εἰσὶ κακῶς διακοινόσαντες καὶ διαρπάσαντες χηρῶν καὶ ὀρφανῶν τὴν ζωήν καὶ έαυτοῖς περιποιησάμενοι ἐκ τῆς διακοινίας ἢς ἔλαβον διακοινήσας εὰν οἱν ἐπιμείνωσι τῇ αὐτῇ ἐπιθυμίᾳ ἁπέθανον καὶ οὐδεμία αὐτοῖς ἐλπίς ζωῆς εὰν δὲ ἐπιστρέψωσι καὶ ἀγνώσ τελείωσωσι τὴν διακοινίαν αὐτῶν δινησόνται ζήσαι. (S9.26.2)

The fate of those who minister badly is stated in stark terms and shows the depth of Hermas' feelings about them, 'If they remain in covetousness they are dead and they have no hope of life'.

There can be no doubt about the fact that Hermas and some the leaders were not in agreement.

Church and World

The relationship of the Christian with the world is an important theme for Hermas. Conflict between the church and the pagan world is inevitable. He uses a motif that is common in wisdom literature, that of a refining-by-fire, in order to encourage those within the church to persevere:

τὸ δὲ χρυσοῖν μέρος ύμείς ἐστε οἱ ἐκφυγόντες τὸν κόσμον τούτον, ὦσπερ γάρ τὸ χρυσὸν δοκιμάζεται διὰ τοῦ πυρᾶ καὶ εὔχρηστον γίνεται οὕτως καὶ ύμείς δοκιμάζεσθε οἱ κατοικοῦντες ἐν αὐτοῖς. οἱ οὖν μείναντες καὶ πυρωθέντες ὕπ᾽ αὐτῶν καθαρισθήσεσθε, ὦσπερ τὸ χρυσὸν ἀποβάλλει τὴν σκωρίαν αὐτοῦ οὕτω καὶ ύμείς ἀποβαλεῖτε πάσαν λύπην καὶ στενοχωρίαν καὶ καθαρισθήσεσθε καὶ χρήσιμοι ἔσεσθε εἰς τὴν οἰκοδομήν τοῦ πύργου. (V4.3.4)

Hermas reveals his stance with regard to the difficulties faced by the Christian community living under Roman government by his use of future verbs, καθαρισθήσεσθε, ἀποβαλεῖτε, ἔσεσθε. The condition that existed for the
Christians was “you who live among them are being tried”, whereas Hermas’ use of the future verbs “you shall be purified”, “you will put away grief”, “you will become useful” are predictive (using Biber’s terminology) and display a confident and defiant stance.

There are further indications of the difficulties faced by the Christian community in S9.28.1-8. This section is an explanation of the eleventh mountain (see S9.1.10) that was heavily wooded with fruit trees. In the explanation the trees bearing the most fruit are a symbol of those who have ‘suffered for the name of the Son of God’ (οἱ παθόντες ὑπὲρ τοῦ δυνάματος τοῦ νιὸν τοῦ θεοῦ) and gave up their lives. Others when ‘brought under authority’ doubted whether to confess or deny, these are represented by trees displaying inferior fruit. Hermas seems to have in mind a past event when Christians suffered persecution, even to death. Osiek suggests that “The situation of such suffering may be a past remembrance, perhaps of the bloody suffering under Nero.”172 Importantly, Hermas turns his attention in S9.28.5 to those who are currently suffering173 for the name. He urges them to see this situation as a blessing, and when they are brought under authority not to hesitate as to denial or confession. Whatever this current persecution might be, Hermas is one who calls for a confident stand of confession.

Whereas conflict with the pagan world was inevitable, Hermas finds that some Christians within the church are acting in a worldly manner and causing conflict in the church. These are likened to half dry and cracked sticks:

{o} δὲ ἡμιζήρους ἔχοντες καὶ ἐν αὐταῖς σχισμᾶς ὁτίοι καὶ διψυχοι καὶ κατάλαλοι εἰς καὶ μηδέποτε εἰρηκούντες εἰς ἑαυτοὺς ἀλλά διεχοστατοῦντες πάντοτε ἀλλὰ καὶ τούτοις φησίν ἐπίκειται μετάνοια. (S8.7.2)

These persons seem to be trouble-makers who are so adept at their task that they cannot live at peace with themselves. They may regard themselves as clever debaters because of their many arguments, but Hermas regards such people as foolish (S8.7.4). They are trying to make a reputation for themselves and desire the first place. Unless they repent and desist from their schisms they will know only death.

Hermas was also in conflict with those who used their wealth to live in luxury and showed no concern for the poor. Hermas considers that such luxury as ‘needless food and drink’ is foolishness (M12.2.1). He warns such people:

172 Osiek, C. The Shepherd of Hermas. p.251
173 Although in V2.2.8 Hermas warns of a great persecution that is coming ie. is yet future. Hermas and the Christian community live under the constant threat of persecution.
There is necessity, or obligation, on these people to refrain from what is essentially, wicked desire. In S1.1.8, Hermas urges those who are rich to look after the widows and orphans because this wealth is ‘beautiful and joyful’.

Conflict grammar also occurs in M10.3.1-4, which deals with the mournful man and the need for joy. Osiek\textsuperscript{174} writes, “The third chapter is a rather straightforward teaching on the quality of joy and the destructiveness of sadness”. No one could doubt that she is correct in this assessment, but there is conflict here for Hermas. Is there more to this than the straightforward teaching that Osiek suggests? In M4.2.3, Hermas says of the ill tempered man that he is grieved at the act which he did and repents. Then in verse 4 he says,

\begin{quote}
αὐτὴ οὖν ἡ λύπη δοκεῖ σωτηρίαν ἔχειν ὀτι τὸ πονη ῥόν πράξας μετενόησεν.
\end{quote}

Did some in the church use grief as a show of repentance and sincerity? Hermas considers mournfulness to be a grieving of the Holy Spirit and calls it ‘wicked grief’.

**Conclusion**

It has been possible to find conflict grammar within a Greek text. However, the results are not so convincing as that of the previous chapter involving the conflict language in 1 Corinthians. This may be because the amount of conflict material in *The Shepherd* is not as great as in 1 Corinthians, where there appeared to be so much material packed into a shorter text. A Greek conflict grammar may require further evaluation, but these preliminary efforts indicate that it should be possible to develop software capable of doing the task.

This chapter has highlighted the areas of conflict Hermas experienced. It was not an exhaustive study because it was limited to the conflict Hermas had within his own church, or with the churches of Rome. Consequently we have not considered Hermas’ self-conflict, or the difficulties he had with his own family.

The analysis has shown that Hermas \textit{was} in conflict with the churches in Rome, because of a changing attitude in the churches. Some Christians were more concerned about wealth and status than caring for others, and this criticism could be brought against the leaders too. He also appears to feel that personal holiness is lacking in the members of the churches beyond his own community. Hermas

\textsuperscript{174} Osiek, C. *The Shepherd of Hermas*. p.138.
compares the present situation with the past, and how it used to be. Although his community seems to be a caring one, and perhaps consisted of poorer members, it seems to be resisting the changes of the age. As Hermas seeks to protect his community from the unwelcome changes, so they become a more insular group turning in on themselves. The life-style that Hermas wants the community to embrace was spirit-oriented and perhaps more rigorous than others. It was already, at the time Hermas wrote, a community in crisis. The next chapter places Hermas’ community into the wider context of the city of Rome.
Chapter 13
PLACING HERMAS IN THE HISTORY OF
CHRISTIANITY IN ROME.

Rome In The First Two Centuries AD.

Roman state and society underwent a major change when the Republic ended and the Imperial monarchy of Augustus’ rule began (27BC). The change in political structure brought a change in the whole social structure of Roman society, a new social class of aristocracy was formed, but the changes took place slowly over 150 years. This same period was marked by an up-turn in the Roman economy fostered by the Pax Romana and increasing urbanisation. Mining and mass production (eg. pottery) being notable examples of areas of economic growth. Demand began to outstrip supply and there were many parts of the Empire which benefited from trade, particularly Spain and Gaul which became centres of production for the Empire. With a rise in urbanisation all over the Empire, there came a rise in demand for food. Soldiers who had settled in the new urban centres demanded olive-oil, meat, fruit, wine and luxury items such as perfume. These were shipped to the remote extremes of the Empire, one of the more usual packing containers for such products being the amphora, and it is the archaeological remains of broken amphora pottery that attests to the bulk and widespread trade that existed.

One of the consequences of this economic boom was the rise in land prices. This occurred in Italy, but parts of Asia were also affected as The Seer remarks:

χοίνιξ σίτου δηναρίου καὶ τρεῖς χοίνικες κριθῶν
dηναρίου, καὶ τὸ ἐλαιον καὶ τὸν οἶνον μὴ ἀδικήσῃς

(Rev 6:6)

This may be a reference to Domitian’s demand that the Asian landowners transfer land from vine-growing to wheat and barley in order to help satisfy demand in Italy.

By the latter part of the 1st century, the most important social division was measured, not so much by money, but by land ownership. The upper strata of

175 MacMullen, R. Roman Social Relations.
177 See Hemer, C.J. The letters to the Seven Churches.
society consisted of the elites in the cities and the rich landowners. The lower strata of society consisted of the poor, freedmen and slaves. In the larger cities, where there were better prospects for employment and entertainment than the countryside could offer, life was still hard and beggars were a common sight in Rome.

The citizens with wealth spent their time in the many taverns that dotted the city, and it was usual that following feasts there were rooms set aside for drinking. Sexual promiscuity was also widespread and may have been linked with drunkenness. Hermas was right to fear that moral laxity might find its way into the church.

Providing help for the poor was still the norm for the Christian communities, despite the few who abused their wealth. Justin writes:

_The wealthy and the willing each gives what he wants as each sees fit, and what is collected is deposited with the presidents. He helps orphans and foreigners staying with us; in a word, he takes care of everyone in distress._

Perhaps things were not as dire as Hermas suggests, but the dangers were always present and he felt it his duty to inform the church of the abuses that did exist.

When economic growth began to slow, the Pax Romana was tested in parts of the Empire and open unrest was a concern for Marcus Aurelius (161-180 AD). Cassius Dio regarded the death of Marcus Aurelius as the end of the golden age of stability (71.36.4). By about 250 AD the reversal of fortunes was complete, the mines were exhausted, food was short, production down and inflation up.

_The Shepherd_ obviously fits into the earlier context well. There were those in the church that Hermas has criticised (chapter 11) for their wealth, acquired through business. Such criticism was also found as a theme (chapter 10) running through the text. This, then, was something that concerned Hermas greatly. Lampe states that “really rich people became discernable for the first time in the 90’s. In the second century the group of those socially elevated increases constantly”. It was this latter period that proved to be the most demanding time for Hermas and his group, as the creation of the new wealthy citizens changed Roman society outside of the church. Especially difficult was the problem of integration of the new wealthy with the

---

178 Philo _Vita 5_. Quoted in Oakes, P. (ed.) _Rome in the Bible and Early Church_. p.86
179 Justin Martyr _Apologies_ 1.6.7. Quoted in Noy, D. _Foreigners at Rome_. p.147
180 Cary, E. (tr). _Dio’s Roman History_.
182 Lampe, P. _From Paul to Valentinus_. p.139

236
already established wealthy groups. Social stratification boundaries needed to be continually re-drawn and this problem was mirrored in the church, particularly the integration of the new wealthy into the church. Hermas condemns the desire for social status as lustful desire and foolishness.

To judge from Romans 16:3-15, it is generally thought that Paul knew of several house churches in Rome. Lampe concludes that from the archaeological evidence it appears that “in the first two centuries there were no house-churches in the sense that specific rooms were permanently set aside for worship in secular houses”183, rather the Christians met in rooms that were used in everyday life. Since there were no cultic rooms, Christians could assemble where it was preferred, or where it was possible, and the host would play a prominent role in the house-church. The character of that church would be affected accordingly.

As has been noted in the previous chapters, Hermas has much to say on the ‘correct’ use of wealth in the church. The selfish use of wealth and living in luxury is regarded as foolishness, as far as Hermas is concerned. It was also noted that ‘foolishness’ is explained as a lack of the spirit’s influence in the life of the believer. The ‘correct’ use of wealth was in the support of the poor, and the community as a whole. But, if the wealthy were to support the poor, the poor had a responsibility to pray for the rich:

ο πένης δὲ ἐπιχορηγοῦμενος ὑπὸ τοῦ πλουσίου ἐνυγχάνει τῷ θεῷ εὐχαριστῶν αὐτῷ, ὑπέρ τοῦ διδόντος αὐτῷ κάκεινος ὑπὶ ἐπισπουδάζει περὶ τοῦ πένητος, ἵνα ἀδιάλειπτος γένηται ἐν τῇ ἐωθί αὐτοῦ: οἴδε γὰρ ὅτι ἡ τοῦ πένητος ἐντευξίς προσδεκτή ἐστι καὶ πλουσία πρὸς κύριον. (S2.1.6)

Hermas adopts and encourages a love ethic that Theissen describes as “love patriarchalism”:

Love patriarchalism takes social differences for granted but ameliorates them through an obligation of respect and love, an obligation upon those who are socially stronger. From the weaker are required subordination, fidelity, and esteem.184

Theissen was, however, describing Pauline communities. Acts gives the impression that Paul travelled from city to city converting wealthy householders with whom he stays (eg. Acts 16:14-15). These households then become centres of worship and fellowship for other Christians who listen to Paul’s teaching. These

183 Lampe, P. From Paul to Valentinus. p.368
184 Theissen, G. The Social Setting of Pauline Christianity. pp.107-108
practices seem to be followed by Paul’s disciples also, for Paul knows of several such house groups already active in Rome (Rm 16:1-16). It seems therefore, that the church of Hermas’ time was not significantly different to that which existed when Paul visited Rome. If Paul was in Roman captivity c.63-c.65, then either not much had changed for Hermas’ community because The Shepherd was written very soon afterwards, or because it was written in a time before too much change could develop. This would therefore suggest a date for The Shepherd within the range of c.70 AD (Coleborne’s suggestion of an early date seems plausible) to c.135 AD (Reilings date for the final redaction).

Under the conditions prevailing in Rome at that time (70-135 AD), it is quite possible that Christian groups were not always aware of the existence of one another, or at least were not in close contact in some cases. It seems that Hermas was aware of other groups of Christians in the city of Rome, for he is concerned about lax moral standards in some of those communities. He also seems to have had some interaction with the leaders of other communities, but he criticises them because they would not give him a hearing. Such limited interaction between groups of Christians encouraged factionalism, and that in turn, allowed a wide diversity of teachers to gain a hearing and a following in Rome. Hermas has difficulties with false teachers who have been causing troubles with their foolish doctrines (S8.6.5), perhaps they had even entered his own community. It is possible, then, that some groups, in order to combat the rise in false doctrine, began to form alliances. It is in this context that we find the importance of the Clement figure, who is to circulate the writings of Hermas to the other churches (V2.4.3). Hermas does not imply that Clement has a leading role in the church other than being a common tie between groups. If this Clement is the same as the author of 1 Clement, we can understand his role as one sending the letter to the Corinthians on behalf of the Roman churches, and not as leader of one influential Roman Christian group, but as representative of many communities. Having such influence in speaking on behalf of the Roman communities on an Empire wide scale, would over a period of time, give the appearance of great authority and would eventually, in an institutionalised church, become a church office.

This adds further confirmation to our suggested dating of The Shepherd. If this Clement is the same person as the author of 1 Clement, and if the date of 1 Clement is correct (towards the end of the first century), then we may also infer that
the *Visions*, at least, could have been in circulation early in the second century.

Maier\textsuperscript{185} suggests that the house church, with its system of patronage, was the source of Hermas' problems. Because the Christian groups could meet wherever was appropriate, under the patronage of the host, it is possible that where the wealthy desired the first positions they could invite sympathetic followers to join them in a new house-church. Hermas, who had once been a successful businessman may have become a leading figure of a house-church (though not a presbyter), but his concern for purity and the looking back to the golden age of the apostles, might indicate that the significance of his group was undermined. Hermas and his supporters were fighting a rear-guard action against the new rich business class of rising status. Hermas has already complained that the leaders of other churches will not give him a hearing, so it would seem that Hermas' group is being marginalised in a changing Roman church scene. Hermas and his group were eccentrics – outside the centre/hub of Roman Christianity.

**Conclusion**

When we place Hermas into the affairs of Rome in the second century we can begin to appreciate why Hermas was so concerned about purity. Luxury is also associated with excess and a decline in moral standards, and Hermas warns about the dangers. Osiek was right in pointing out that *The Shepherd* deals with the theme of rich and poor, but it is a subsidiary of the main theme, which is the call to purity.

While the theme of 'rich and poor' has thus been the subject of rigorous treatment by scholars, some of the themes highlighted by the methods of this study have not been given extensive consideration, and probably should be.

\textsuperscript{185} Maier, H.O. *The Social Setting of Ministry*. p.64
Chapter 14

CONCLUSION AND FUTURE RESEARCH

For many years *The Shepherd of Hermas* has been a problematic text for those wishing to trace the development of the church from Jewish-Christian beginnings to an institutionalised church. Tracing the development chronologically from Paul through (possibly Hebrews)/1 Clement and *The Shepherd* would require a reversal of trend when dealing with *The Shepherd*. It does not fit a linear progression, but seems rather, to revert to the early Pauline church model Therefore, advocates of a theory of development find it necessary to date *The Shepherd* earlier than what had become the "traditional" date of c.150 AD. This thesis has confirmed that they are probably right to do so, but it has arrived at this conclusion by a different route.

In order to solve the question of single or multiple authorship, modern multivariate statistics was employed. For the past decade the method of choice in the statistical analysis of texts has been that of PCA. This method was found to be deficient in a number of respects, but particularly with regard to the problem of data choice and data transformation. Slight changes in these choices can produce large and crucial changes in the result obtained, and therefore the conclusions drawn from those results. All of the traditional statistical methods, from Components Analysis to Multidimensional Scaling were shown to have had similar problems associated with them. Random Forest was the method finally chosen to do the task required, because it did not require ANY data selection (other than function word) or any data transformation.

The main problem of the method is that of the choice of the variable used in the analysis, but the algorithm overcomes this by using a bootstrap re-sampling in order to allow an indication of how each variable contributes to, and affects, the final result. A simple majority vote determines the 'best choice' result. Future research should concern itself with relying less on the majority vote feature and should replace it with a Kullback-Leibler information algorithm. This, I believe, would give more reliable results. However, using unsupervised random forest (without group labels attached) allows calculation of a similarity matrix, the proximity matrix, which can then be used in any hierarchical clustering method or in scaling methods such as MDS to assign cases to clusters. Good results were obtained from random forest and
PAM clustering

Using the methods has shown that the single author hypothesis is probably the correct one. This was a significant result, not only for those interested in The Shepherd for its own sake, but also for those interested in the social-scientific understanding of the early Christian communities, because now, for the first time, we are able to assign The Shepherd to a more suitable date. The date could only be assigned when the single author hypothesis had been proved and the internal evidence obtained by an analysis of the content of the text. This thesis chose to look at the MAJOR themes of The Shepherd, since this would indicate the sort of things that Hermas thought were important for the church to know, and to look at the conflict that Hermas encountered from other churches.

The section dealing with the themes in The Shepherd required the development, both theoretical and practical, of a method of analysis based on lexical repetition. The idea was first tested on Vision 1 and proved to be so successful that it was then used to show theme development through the whole of The Shepherd. This showed that Hermas was particularly concerned with the purity of the church. He elaborated on this theme showing the need for repentance, the foolishness of luxury, and commitment to 'love patriarchalism'. He calls for a return to the morals and standards of a past age, remembers the time of the apostles, a time which he fondly recalls. If these apostles were Peter and Paul (for both have a connection with Rome), it gives us an indication that The Shepherd requires an earlier rather than a later date. We are now in the position to understand the concerns of the Hermas community.

The section on conflict language called for the development of methodology never before used on Greek texts. It was tested on 1 Corinthians, where it proved that it had the potential of producing good results. When used to determine the conflict that Hermas had with other Christian communities, some surprising instances of conflict were noted. In particular, it suggests that Hermas WAS in conflict with some of the other Christian communities in Rome. In these conflicts Hermas was marginalized and this is suggestive of the fact that Hermas' community was already in decline, lacking the ability to inform or change the practice of the wider church. Hermas was probably opposed to the institutionalisation of the church in a fashion informed by the values of the Empire and the experience of the business classes since he took his model of the church from a past age. The Shepherd was,
perhaps, Hermas’ last attempt to get a hearing and inform the church. In a way he succeeded because The Shepherd was used as an aid to catechesis in many churches long after Hermas.

The authorship analysis has shown that this was Hermas’ own work. He has been criticised in the past for his awkward use of language and for his confusion of theme. This work has shown the opposite to be true of Hermas. He wrote and edited his own work, which indicates an above average ability. Granted, there are constructions which display awkwardness of language and confusion of theme explanation (if given at all). This might have more to do with the way the material was presented than in the literary ability of the author. It was hinted that some of the material may contain residue of an originally oral work. This thesis has not investigated the question of orality, or that the text may be in the style of an ancient sermon. These questions are for future research, but the methods to achieve this end have been provided in this research.

Was Hermas a “typical” representative of the church in Rome? Osiek agrees with Lampe in the following assessment, that Hermas lacks the education and ability to be a scholar of the church, and can best be summed up as “typical” of the many, the “ordinary” Christian. This thesis asserts that Hermas is NOT a typical representative of the church in that respect however.

The conflict he experiences suggests that Hermas was one of a dying breed, already stuck in the past, old fashioned. He shows all the elements of coming from a Jewish-Christian background, though influenced by Hellenism in the past. He knows of, and uses in modified form, the ideas of Hellenistic pagan literature. Although no evidence has suggested itself in this thesis, there are some who see influences of Qumran in The Shepherd. This would be an item for future research, not least given that purity has emerged as a key theme in the work.

More importantly, the methods were designed to be general enough to use on ANY Greek text. The methodology used could be applied to any Greek text without additional refinement. In the age of machine learning, these methods could be programmed to extract information from texts with limited operator involvement, thus removing some of the subjectivity injected into textual studies by modern criticism (see Wilson in chapter one of this thesis), moreover, when applied to a corpus of Greek literature containing varied style and covering a greater range of historical periods, the methods would be ‘sharpened and improved’.

242
Greek literature already exists (Perseus Project) but is not in a suitable format for these studies.

The following topics for future research have therefore been identified:

i. The Random Forest algorithm uses a system of majority voting in order to assign cases. This algorithm should be written so that it compares probability density using the Kullback-Liebler function. This would improve the performance and reliability of the Random Forest method.

ii. The Shepherd should be analysed for evidence of orality and sermon style.

iii. The method for determining conflict grammar should be refined by using corpus based texts, allowing the use of, and the change in, conflict language from different literature types and periods.

iv. The Shepherd should be analysed for any points of comparison between it and the literature and concerns of Qumran community.

Of Hermas himself we can truly say:

τῶν τοιούτων οὖν ἡ πάροδος μετὰ τῶν ἀγγέλων ἐστίν
Appendix 1.1
The Muratorian Fragment

. . . at which nevertheless he was present, and so he placed [them in his narrative]. The third book of the Gospel is that according to Luke. Luke, the well-known physician, after the ascension of Christ, when Paul had taken with him as one zealous for the law, composed it in his own name, according to [the general] belief. Yet he himself had not seen the Lord in the flesh; and therefore, as he was able to ascertain events, so indeed he begins to tell the story from the birth of John. The fourth of the Gospels is that of John, [one] of the disciples. To his fellow disciples and bishops, who had been urging him [to write], he said, 'Fast with me from today to three days, and what will be revealed to each one let us tell it to one another.' In the same night it was revealed to Andrew, [one] of the apostles, that John should write down all things in his own name while all of them should review it. And so, though various elements may be taught in the individual books of the Gospels, nevertheless this makes no difference to the faith of believers, since by the one sovereign Spirit all things have been declared in all [the Gospels]: concerning the nativity, concerning the passion, concerning the resurrection, concerning life with his disciples, and concerning his twofold coming; the first in lowliness when he was despised, which has taken place, the second glorious in royal power, which is still in the future. What marvel is it then, if John so consistently mentions these particular points also in his Epistles, saying about himself, 'What we have seen with our eyes and heard with our ears and our hands have handled, these things we have written to you?' For in this way he professes [himself] to be not only an eyewitness and hearer, but also a writer of all the marvellous deeds of the Lord, in their order. Moreover, the acts of all the apostles were written in one book. For 'most excellent Theophilus' Luke compiled the individual events that took place in his presence as he plainly shows by omitting the martyrdom of Peter as well as the departure of Paul from the city [of Rome] when he journeyed to Spain. As for the Epistles of Paul, they themselves make clear to those desiring to understand, which ones [they are], from what place, or for what reason they were sent. First of all, to the Corinthians, prohibiting their heretical schisms; next, to the Galatians, against circumcision; then to the Romans he wrote at length, explaining the order (or, plan) of the Scriptures, and also that Christ is their principle (or, main theme). It is necessary for us to discuss these one by one, since the blessed apostle Paul himself, following the example of his predecessor John, writes by name to only seven churches in the following sequence: To the Corinthians first, to the Ephesians second, to the Philippians third, to the Colossians fourth, to the Galatians fifth, to the Thessalonians sixth, to the Romans seventh. It is true that he writes once more to the Corinthians and to the Thessalonians for the sake of admonition, yet it is clearly recognizable that there is one Church spread throughout the whole extent of the earth. For John also in the Apocalypse, though he writes to seven churches, nevertheless speaks to all. [Paul also wrote] out of affection and love one to Philemon, one to Titus, and two to Timothy; and these are held sacred in the esteem of the Church catholic for the regulation of ecclesiastical discipline. There is current also [an epistle] to the Laodiceans, [and] another to the Alexandrians, [both] forged in Paul’s name to [further] the heresy of Marcion, and several others which cannot be received into the catholic Church for it is not fitting that gall be mixed with honey. Moreover, the epistle of Jude and two of the above-mentioned (or, bearing the name of) John are counted (or, used) in the catholic [Church]; and [the book of] Wisdom, written by the friends of Solomon in his honour. We receive only the apocalypses of John and Peter, though some of us are not willing that the latter be read in church. But Hermas wrote the Shepherd very recently, in our times, in the city of Rome, while bishop Pius, his brother, was occupying the [episcopal] chair of the church of the city of Rome. And therefore it ought indeed to be read; but it cannot be read publicly to the people in church either among the Prophets, whose number is complete, or among the Apostles, for it is after [their] time. But we accept nothing whatever of Arsinous or Valentinus or Miltiades, who also composed a new book of psalms for Marcion, together with Basilides, the Asian founder of the Cataphrygians . . .
Taken from,
Hans Lietzmann, *Das Muratorische Fragment und die Monarchianischen Prologue zu den Evangelien* (Kleine Texte, i; Bonn, 1902; 2nd ed., Berlin, 1933)
# Appendix 3.1
## Master Set of Function Words

<table>
<thead>
<tr>
<th>Count1</th>
<th>αχρα , αχρασ</th>
<th>Count51</th>
<th>δήμου</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count2</td>
<td>αφω</td>
<td>Count52</td>
<td>δικ , δι '</td>
</tr>
<tr>
<td>Count3</td>
<td>αλλας</td>
<td>Count53</td>
<td>δηνεκας</td>
</tr>
<tr>
<td>Count4</td>
<td>αν</td>
<td>Count54</td>
<td>δικαίος</td>
</tr>
<tr>
<td>Count5</td>
<td>ανεν</td>
<td>Count55</td>
<td>διο</td>
</tr>
<tr>
<td>Count6</td>
<td>αντικρας</td>
<td>Count56</td>
<td>διόπερ</td>
</tr>
<tr>
<td>Count7</td>
<td>ανω , ανωθεν</td>
<td>Count57</td>
<td>διτι</td>
</tr>
<tr>
<td>Count8</td>
<td>αρτι</td>
<td>Count58</td>
<td>δισσως</td>
</tr>
<tr>
<td>Count9</td>
<td>ατερ</td>
<td>Count59</td>
<td>δις</td>
</tr>
<tr>
<td>Count10</td>
<td>αδειας</td>
<td>Count60</td>
<td>διχα</td>
</tr>
<tr>
<td>Count11</td>
<td>αδηλως</td>
<td>Count61</td>
<td>διφατως</td>
</tr>
<tr>
<td>Count12</td>
<td>αδικατπως</td>
<td>Count62</td>
<td>δισκυλως</td>
</tr>
<tr>
<td>Count13</td>
<td>αδικως</td>
<td>Count63</td>
<td>δισφεδν</td>
</tr>
<tr>
<td>Count14</td>
<td>αει</td>
<td>Count64</td>
<td>εβαραστι</td>
</tr>
<tr>
<td>Count15</td>
<td>αφοβως</td>
<td>Count65</td>
<td>εμπροσθε , εμπροσθεν</td>
</tr>
<tr>
<td>Count16</td>
<td>αφοβως</td>
<td>Count66</td>
<td>εναντι</td>
</tr>
<tr>
<td>Count17</td>
<td>αγνως</td>
<td>Count67</td>
<td>ενδοθεν</td>
</tr>
<tr>
<td>Count18</td>
<td>ακεβως</td>
<td>Count68</td>
<td>ενδα</td>
</tr>
<tr>
<td>Count19</td>
<td>ακωλυτως</td>
<td>Count69</td>
<td>ενθεν</td>
</tr>
<tr>
<td>Count20</td>
<td>αληθως</td>
<td>Count70</td>
<td>επείτα</td>
</tr>
<tr>
<td>Count21</td>
<td>αλλαχοι</td>
<td>Count71</td>
<td>εςω , εςωθεν</td>
</tr>
<tr>
<td>Count22</td>
<td>αλλαχθεν</td>
<td>Count72</td>
<td>επι</td>
</tr>
<tr>
<td>Count23</td>
<td>αλλα , αλλ '</td>
<td>Count73</td>
<td>εξω , εξωθεν</td>
</tr>
<tr>
<td>Count24</td>
<td>αμεμπτως</td>
<td>Count74</td>
<td>ειλεκεν</td>
</tr>
<tr>
<td>Count25</td>
<td>αναμηρρητως</td>
<td>Count75</td>
<td>ειτα , ειτεν</td>
</tr>
<tr>
<td>Count26</td>
<td>ανα</td>
<td>Count76</td>
<td>ει , ειπερ , ειπως</td>
</tr>
<tr>
<td>Count27</td>
<td>ανδρειως</td>
<td>Count77</td>
<td>εικη</td>
</tr>
<tr>
<td>Count28</td>
<td>ανδιμως</td>
<td>Count78</td>
<td>εαν , εανπερ</td>
</tr>
<tr>
<td>Count29</td>
<td>αντιπέρα</td>
<td>Count79</td>
<td>εχες</td>
</tr>
<tr>
<td>Count30</td>
<td>αντι , ανθ ' , αντ '</td>
<td>Count80</td>
<td>εφαπαξ</td>
</tr>
<tr>
<td>Count31</td>
<td>απεναντι</td>
<td>Count81</td>
<td>εγγυος</td>
</tr>
<tr>
<td>Count32</td>
<td>απλως</td>
<td>Count82</td>
<td>εκ , εξ</td>
</tr>
<tr>
<td>Count33</td>
<td>αποτομως</td>
<td>Count83</td>
<td>εκει , εκειθεν</td>
</tr>
<tr>
<td>Count34</td>
<td>απο , απ ' , αφ '</td>
<td>Count84</td>
<td>εκείσε</td>
</tr>
<tr>
<td>Count35</td>
<td>απροσωπολημπτως</td>
<td>Count85</td>
<td>εκουσως</td>
</tr>
<tr>
<td>Count36</td>
<td>ασφαλως</td>
<td>Count86</td>
<td>εκτενως</td>
</tr>
<tr>
<td>Count37</td>
<td>ασφαλως</td>
<td>Count87</td>
<td>εναλλαζ</td>
</tr>
<tr>
<td>Count38</td>
<td>ασφαλως</td>
<td>Count88</td>
<td>ενδοξως</td>
</tr>
<tr>
<td>Count39</td>
<td>αυρων</td>
<td>Count89</td>
<td>ενιοτε</td>
</tr>
<tr>
<td>Count40</td>
<td>αμα</td>
<td>Count90</td>
<td>ενιοδε</td>
</tr>
<tr>
<td>Count101</td>
<td>επεκεινα</td>
<td>Count102</td>
<td>επιμελως</td>
</tr>
<tr>
<td>Count103</td>
<td>επι , επ ' , εφ '</td>
<td>Count104</td>
<td>ετοιμως</td>
</tr>
<tr>
<td>Count105</td>
<td>εξαφωνης</td>
<td>Count106</td>
<td>εξαπωνα</td>
</tr>
<tr>
<td>Count107</td>
<td>εξης</td>
<td>Count108</td>
<td>ευ</td>
</tr>
<tr>
<td>Count109</td>
<td>ευγε</td>
<td>Count110</td>
<td>ευαρεστως</td>
</tr>
<tr>
<td>Count111</td>
<td>εωιαιρως</td>
<td>Count112</td>
<td>ευπερως</td>
</tr>
<tr>
<td>Count113</td>
<td>ειδθεως</td>
<td>Count114</td>
<td>εινης</td>
</tr>
<tr>
<td>Count115</td>
<td>εινθωμως</td>
<td>Count116</td>
<td>εινθωμως</td>
</tr>
<tr>
<td>Count117</td>
<td>εινθωμως</td>
<td>Count118</td>
<td>εινθωμως</td>
</tr>
<tr>
<td>Count119</td>
<td>εινεκα</td>
<td>Count120</td>
<td>εινεκεν</td>
</tr>
<tr>
<td>Count121</td>
<td>ειως</td>
<td>Count122</td>
<td>εινεκρως</td>
</tr>
<tr>
<td>Count123</td>
<td>εινεκρως</td>
<td>Count124</td>
<td>εινεκρως</td>
</tr>
<tr>
<td>Count125</td>
<td>φρονιμως</td>
<td>Count126</td>
<td>γαρ</td>
</tr>
<tr>
<td>Count127</td>
<td>γε</td>
<td>Count128</td>
<td>γησιως</td>
</tr>
<tr>
<td>Count129</td>
<td>γοιν</td>
<td>Count130</td>
<td>γοινως</td>
</tr>
<tr>
<td>Count131</td>
<td>γινικα</td>
<td>Count132</td>
<td>γη , γηοι</td>
</tr>
<tr>
<td>Count133</td>
<td>γηδη</td>
<td>Count134</td>
<td>γη , γηπερ</td>
</tr>
<tr>
<td>Count135</td>
<td>γινατι</td>
<td>Count136</td>
<td>γαγως</td>
</tr>
<tr>
<td>Count137</td>
<td>γαγως</td>
<td>Count138</td>
<td>γων , γυ '</td>
</tr>
<tr>
<td>Count139</td>
<td>γαν</td>
<td>Count140</td>
<td>γαι</td>
</tr>
</tbody>
</table>
## Appendix 6.1

The first six component loadings of a PCA analysis.

<table>
<thead>
<tr>
<th></th>
<th>Comp.1</th>
<th>Comp.2</th>
<th>Comp.3</th>
<th>Comp.4</th>
<th>Comp.5</th>
<th>Comp.6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count111</td>
<td>0.274</td>
<td>-0.165</td>
<td>0.019</td>
<td>0.031</td>
<td>0.005</td>
<td>0.029</td>
</tr>
<tr>
<td>Count37</td>
<td>-0.066</td>
<td>0.292</td>
<td>0.129</td>
<td>-0.283</td>
<td>-0.019</td>
<td>-0.053</td>
</tr>
<tr>
<td>Count147</td>
<td>-0.121</td>
<td>-0.244</td>
<td>-0.189</td>
<td>-0.096</td>
<td>-0.192</td>
<td>-0.007</td>
</tr>
<tr>
<td>Count63</td>
<td>-0.018</td>
<td>0.074</td>
<td>0.029</td>
<td>0.291</td>
<td>0.105</td>
<td>0.319</td>
</tr>
<tr>
<td>Count97</td>
<td>-0.254</td>
<td>-0.081</td>
<td>-0.027</td>
<td>-0.179</td>
<td>0.046</td>
<td>0.114</td>
</tr>
<tr>
<td>Count211</td>
<td>0.068</td>
<td>0.260</td>
<td>0.191</td>
<td>-0.087</td>
<td>-0.119</td>
<td>0.070</td>
</tr>
<tr>
<td>Count137</td>
<td>0.112</td>
<td>0.143</td>
<td>-0.017</td>
<td>-0.141</td>
<td>-0.207</td>
<td>-0.065</td>
</tr>
<tr>
<td>Count39</td>
<td>-0.245</td>
<td>-0.019</td>
<td>0.167</td>
<td>0.056</td>
<td>-0.001</td>
<td>0.031</td>
</tr>
<tr>
<td>Count109</td>
<td>-0.208</td>
<td>0.104</td>
<td>-0.057</td>
<td>0.285</td>
<td>-0.161</td>
<td>0.117</td>
</tr>
<tr>
<td>Count238</td>
<td>-0.084</td>
<td>-0.325</td>
<td>-0.025</td>
<td>0.065</td>
<td>-0.019</td>
<td>0.073</td>
</tr>
<tr>
<td>Count18</td>
<td>-0.285</td>
<td>0.058</td>
<td>0.036</td>
<td>0.117</td>
<td>-0.098</td>
<td>0.056</td>
</tr>
<tr>
<td>Count60</td>
<td>-0.090</td>
<td>0.048</td>
<td>-0.286</td>
<td>-0.206</td>
<td>-0.121</td>
<td>0.139</td>
</tr>
<tr>
<td>Count124</td>
<td>-0.081</td>
<td>-0.141</td>
<td>0.293</td>
<td>-0.136</td>
<td>-0.047</td>
<td>0.067</td>
</tr>
<tr>
<td>Count59</td>
<td>-0.247</td>
<td>0.062</td>
<td>0.044</td>
<td>-0.021</td>
<td>-0.033</td>
<td>0.122</td>
</tr>
<tr>
<td>Count195</td>
<td>-0.052</td>
<td>0.061</td>
<td>-0.166</td>
<td>-0.014</td>
<td>-0.050</td>
<td>-0.380</td>
</tr>
<tr>
<td>Count103</td>
<td>-0.188</td>
<td>-0.036</td>
<td>-0.073</td>
<td>-0.251</td>
<td>0.083</td>
<td>0.242</td>
</tr>
<tr>
<td>Count178</td>
<td>-0.148</td>
<td>-0.213</td>
<td>-0.098</td>
<td>-0.159</td>
<td>-0.174</td>
<td>-0.064</td>
</tr>
<tr>
<td>Count187</td>
<td>-0.027</td>
<td>0.095</td>
<td>-0.264</td>
<td>-0.126</td>
<td>-0.157</td>
<td>0.014</td>
</tr>
<tr>
<td>Count236</td>
<td>-0.211</td>
<td>0.004</td>
<td>0.064</td>
<td>-0.193</td>
<td>0.144</td>
<td>-0.080</td>
</tr>
<tr>
<td>Count157</td>
<td>-0.123</td>
<td>0.167</td>
<td>0.188</td>
<td>0.029</td>
<td>-0.202</td>
<td>-0.122</td>
</tr>
<tr>
<td>Count191</td>
<td>-0.073</td>
<td>0.030</td>
<td>0.093</td>
<td>-0.199</td>
<td>0.021</td>
<td>-0.147</td>
</tr>
<tr>
<td>Count4</td>
<td>0.001</td>
<td>0.196</td>
<td>-0.146</td>
<td>-0.155</td>
<td>-0.116</td>
<td>0.204</td>
</tr>
<tr>
<td>Count142</td>
<td>-0.216</td>
<td>0.077</td>
<td>0.045</td>
<td>-0.156</td>
<td>0.198</td>
<td>-0.158</td>
</tr>
<tr>
<td>Count64</td>
<td>0.159</td>
<td>0.208</td>
<td>0.017</td>
<td>-0.080</td>
<td>-0.188</td>
<td>0.127</td>
</tr>
<tr>
<td>Count93</td>
<td>0.116</td>
<td>0.234</td>
<td>0.163</td>
<td>-0.202</td>
<td>-0.085</td>
<td>0.186</td>
</tr>
<tr>
<td>Count57</td>
<td>0.042</td>
<td>0.084</td>
<td>0.042</td>
<td>-0.027</td>
<td>-0.085</td>
<td>0.065</td>
</tr>
<tr>
<td>Count123</td>
<td>-0.240</td>
<td>0.058</td>
<td>0.085</td>
<td>0.172</td>
<td>-0.174</td>
<td>-0.055</td>
</tr>
<tr>
<td>Count173</td>
<td>-0.122</td>
<td>-0.001</td>
<td>-0.127</td>
<td>-0.055</td>
<td>0.211</td>
<td>0.266</td>
</tr>
<tr>
<td>Count233</td>
<td>-0.191</td>
<td>-0.197</td>
<td>0.117</td>
<td>0.095</td>
<td>-0.234</td>
<td>-0.022</td>
</tr>
<tr>
<td>Count190</td>
<td>-0.131</td>
<td>0.169</td>
<td>-0.180</td>
<td>0.197</td>
<td>-0.126</td>
<td>-0.092</td>
</tr>
<tr>
<td>Count184</td>
<td>-0.064</td>
<td>0.139</td>
<td>-0.314</td>
<td>-0.006</td>
<td>0.025</td>
<td>-0.013</td>
</tr>
<tr>
<td>Count73</td>
<td>0.086</td>
<td>-0.135</td>
<td>-0.031</td>
<td>0.209</td>
<td>0.172</td>
<td>0.140</td>
</tr>
<tr>
<td>Count214</td>
<td>-0.201</td>
<td>0.157</td>
<td>-0.219</td>
<td>-0.053</td>
<td>0.092</td>
<td>-0.006</td>
</tr>
<tr>
<td>Count171</td>
<td>-0.150</td>
<td>-0.026</td>
<td>-0.110</td>
<td>-0.083</td>
<td>0.273</td>
<td>0.254</td>
</tr>
<tr>
<td>Count223</td>
<td>-0.073</td>
<td>0.057</td>
<td>0.312</td>
<td>-0.039</td>
<td>0.173</td>
<td>-0.168</td>
</tr>
<tr>
<td>Count21</td>
<td>0.134</td>
<td>0.021</td>
<td>0.174</td>
<td>-0.131</td>
<td>-0.170</td>
<td>0.174</td>
</tr>
<tr>
<td></td>
<td>Count242</td>
<td>Count230</td>
<td>Count210</td>
<td>Count33</td>
<td>Count114</td>
<td>Count58</td>
</tr>
<tr>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>---------</td>
<td>----------</td>
<td>---------</td>
</tr>
<tr>
<td></td>
<td>-0.139</td>
<td>-0.044</td>
<td>0.037</td>
<td>-0.068</td>
<td>-0.132</td>
<td>0.015</td>
</tr>
<tr>
<td></td>
<td>0.026</td>
<td>0.070</td>
<td>-0.077</td>
<td>-0.239</td>
<td>0.199</td>
<td>0.234</td>
</tr>
<tr>
<td></td>
<td>0.215</td>
<td>-0.028</td>
<td>-0.066</td>
<td>0.097</td>
<td>-0.093</td>
<td>-0.028</td>
</tr>
<tr>
<td></td>
<td>-0.018</td>
<td>-0.033</td>
<td>-0.129</td>
<td>-0.056</td>
<td>0.301</td>
<td>0.184</td>
</tr>
<tr>
<td></td>
<td>0.005</td>
<td>0.163</td>
<td>-0.130</td>
<td>-0.315</td>
<td>-0.153</td>
<td>-0.069</td>
</tr>
<tr>
<td></td>
<td>0.103</td>
<td>-0.268</td>
<td>0.172</td>
<td>0.038</td>
<td>-0.014</td>
<td>-0.009</td>
</tr>
</tbody>
</table>
Appendix 7.1

Vector of variable weights obtained by OVW

<p>| 1 ἀχρι , ἀχρις | 0.00036 | 141 μηδέ , μηδ’ | 0.00188 |
| 2 ἀφνω | 0.00005 | 142 μηδέποτε | 0.00000 |
| 3 ἀλλως | 0.00786 | 143 μή | 0.00001 |
| 4 ἀν | 0.00119 | 144 μήν | 0.00021 |
| 5 ἀνευ | 0.00488 | 145 μήποτε | 0.00057 |
| 6 ἀνω , ἀνωθεν | 0.00357 | 146 μήτε | 0.00128 |
| 7 ἀρτι | 0.00513 | 147 μήτι | 0.01100 |
| 8 ἀτερ | 0.01568 | 148 μόλις | 0.00104 |
| 9 ἀδιαλειπτως | 0.01108 | 149 ναὶ | 0.00003 |
| 10 ἀδίκως | 0.00001 | 150 νή | 0.00526 |
| 11 ἀδί | 0.00536 | 151 νῦν , νῦν | 0.00001 |
| 12 ἀφόβως | 0.00348 | 152 όμοίως | 0.00225 |
| 13 ἀφόνως | 0.00898 | 153 όμολογομένως | 0.00058 |
| 14 ἀγνώς | 0.00274 | 154 όμοθυμαδόν | 0.00883 |
| 15 ἀκριβῶς | 0.01493 | 155 όμοῦ | 0.01372 |
| 16 ἀληθῶς | 0.00780 | 156 όσάκης | 0.00559 |
| 17 ἀλλαχόθεν | 0.00895 | 157 όσίως | 0.01554 |
| 18 ἀλλά , ἀλλ’ | 0.00000 | 158 ὀπισθε , ὀπισθεν | 0.00031 |
| 19 ἀμέμπτως | 0.00788 | 159 ὀπίσω | 0.00664 |
| 20 ἀναντιρήτως | 0.01755 | 160 ὄργιλως | 0.00879 |
| 21 ἀνά | 0.00007 | 161 ὄρθως | 0.00841 |
| 22 ἀνδρέως | 0.00288 | 162 ὄφε | 0.00004 |
| 23 ἀνόμως | 0.00009 | 163 οὐπω | 0.00076 |
| 24 ἀντί , ἀνθ’ , ἀντ’ | 0.00007 | 164 οὐτε | 0.00058 |
| 25 ἀπέναντι | 0.00260 | 165 οὐ , οὐχ , οὐκ | 0.00000 |
| 26 ἀπλῶς | 0.00824 | 166 οὐχί | 0.00139 |
| 27 ἀποτάμως | 0.00172 | 167 οὐδέ | 0.00230 |
| 28 ἀπό , ἀπ’ , ἀφ’ | 0.00000 | 168 οὐδέποτε | 0.00338 |
| 29 ἀφαλῶς | 0.00001 | 169 οὐδέπω | 0.00086 |
| 30 ἀφριων | 0.00114 | 170 οὐκέτι | 0.00110 |
| 31 ἀμα | 0.00806 | 171 οὐτω , οὔτως | 0.00006 |
| 32 ἀπαξ | 0.01243 | 172 ὄλως | 0.00040 |
| 33 ἄχαμαί | 0.00243 | 173 ὄμως | 0.00703 |
| 34 ἄριν | 0.00224 | 174 ὄπου | 0.00701 |
| 35 χωρίς | 0.00239 | 175 ὄπως | 0.00834 |</p>
<table>
<thead>
<tr>
<th>Λόγος</th>
<th>Βάρος</th>
</tr>
</thead>
<tbody>
<tr>
<td>36 δεινώς</td>
<td>0.00803</td>
</tr>
<tr>
<td>37 δεύρο</td>
<td>0.00004</td>
</tr>
<tr>
<td>38 δέ</td>
<td>0.00000</td>
</tr>
<tr>
<td>39 δή</td>
<td>0.00710</td>
</tr>
<tr>
<td>40 δία , δί</td>
<td>0.00000</td>
</tr>
<tr>
<td>41 δικαίως</td>
<td>0.00291</td>
</tr>
<tr>
<td>42 δίο</td>
<td>0.01385</td>
</tr>
<tr>
<td>43 διόπερ</td>
<td>0.00602</td>
</tr>
<tr>
<td>44 διότι</td>
<td>0.00700</td>
</tr>
<tr>
<td>45 διστώς</td>
<td>0.00971</td>
</tr>
<tr>
<td>46 δίς</td>
<td>0.01468</td>
</tr>
<tr>
<td>47 δυνάτως</td>
<td>0.01493</td>
</tr>
<tr>
<td>48 δυσκόλως</td>
<td>0.00031</td>
</tr>
<tr>
<td>49 δωρεάν</td>
<td>0.00195</td>
</tr>
<tr>
<td>50 ἐβραϊστί</td>
<td>0.00003</td>
</tr>
<tr>
<td>51 ἐμπροσθε , ἐμπροσθεν</td>
<td>0.00813</td>
</tr>
<tr>
<td>52 ἐναυτι</td>
<td>0.00324</td>
</tr>
<tr>
<td>53 ἐνδοθεν</td>
<td>0.00722</td>
</tr>
<tr>
<td>54 ἐνθεν</td>
<td>0.01668</td>
</tr>
<tr>
<td>55 ἐπείτα</td>
<td>0.00850</td>
</tr>
<tr>
<td>56 ἐσω , ἐσωθεν</td>
<td>0.00004</td>
</tr>
<tr>
<td>57 ἐτι</td>
<td>0.00291</td>
</tr>
<tr>
<td>58 ἐξω , ἐξωθεν</td>
<td>0.00113</td>
</tr>
<tr>
<td>59 ει , ειπερ , ειπως</td>
<td>0.00011</td>
</tr>
<tr>
<td>60 εάν , εάνπερ</td>
<td>0.00000</td>
</tr>
<tr>
<td>61 ἐφάπαξ</td>
<td>0.00383</td>
</tr>
<tr>
<td>62 ἐγγύς</td>
<td>0.01411</td>
</tr>
<tr>
<td>63 ἐκ , ἐξ</td>
<td>0.00000</td>
</tr>
<tr>
<td>64 ἐκεῖ , ἐκεῖθεν</td>
<td>0.00003</td>
</tr>
<tr>
<td>65 ἐκείσε</td>
<td>0.00236</td>
</tr>
<tr>
<td>66 ἐκουσίως</td>
<td>0.01136</td>
</tr>
<tr>
<td>67 ἐνδόξως</td>
<td>0.00665</td>
</tr>
<tr>
<td>68 ἐνθάδε</td>
<td>0.00456</td>
</tr>
<tr>
<td>69 ἐνταῦθα</td>
<td>0.00088</td>
</tr>
<tr>
<td>70 ἐντείθεν</td>
<td>0.00163</td>
</tr>
<tr>
<td>71 ἐντός</td>
<td>0.00179</td>
</tr>
<tr>
<td>72 ἐνώπιον</td>
<td>0.00072</td>
</tr>
<tr>
<td>73 ἐπαύριον</td>
<td>0.00049</td>
</tr>
<tr>
<td>74 ἐπάν</td>
<td>0.00926</td>
</tr>
<tr>
<td>75 ἐπάνω</td>
<td>0.00000</td>
</tr>
<tr>
<td>76 ἐπειδή</td>
<td>0.01343</td>
</tr>
<tr>
<td>176 θεῖν</td>
<td>0.000520</td>
</tr>
<tr>
<td>177 θαν</td>
<td>0.00110</td>
</tr>
<tr>
<td>178 οτι</td>
<td>0.00001</td>
</tr>
<tr>
<td>179 πανταχού</td>
<td>0.00560</td>
</tr>
<tr>
<td>180 παραχρήμα</td>
<td>0.00757</td>
</tr>
<tr>
<td>181 παρά , παρ ′</td>
<td>0.00005</td>
</tr>
<tr>
<td>182 παρεκτός</td>
<td>0.00694</td>
</tr>
<tr>
<td>183 πάλαι</td>
<td>0.00071</td>
</tr>
<tr>
<td>184 πάλιν</td>
<td>0.00005</td>
</tr>
<tr>
<td>185 πάντως</td>
<td>0.00000</td>
</tr>
<tr>
<td>186 περισσοτέρως</td>
<td>0.00909</td>
</tr>
<tr>
<td>187 περί</td>
<td>0.00001</td>
</tr>
<tr>
<td>188 πέραν</td>
<td>0.01204</td>
</tr>
<tr>
<td>189 πέρυσι , πέρυσιν</td>
<td>0.00227</td>
</tr>
<tr>
<td>190 πικρῶς</td>
<td>0.00031</td>
</tr>
<tr>
<td>191 πλήν</td>
<td>0.00686</td>
</tr>
<tr>
<td>192 πλούσιως</td>
<td>0.00513</td>
</tr>
<tr>
<td>193 πνευματικῶς</td>
<td>0.01503</td>
</tr>
<tr>
<td>194 πολλάκις</td>
<td>0.00064</td>
</tr>
<tr>
<td>195 ποσάκις</td>
<td>0.01699</td>
</tr>
<tr>
<td>196 ποτέ</td>
<td>0.00002</td>
</tr>
<tr>
<td>197 πόθεν</td>
<td>0.00271</td>
</tr>
<tr>
<td>198 πόρρω , πόρρωθεν</td>
<td>0.00002</td>
</tr>
<tr>
<td>199 πότερον</td>
<td>0.00029</td>
</tr>
<tr>
<td>200 πρίν</td>
<td>0.00567</td>
</tr>
<tr>
<td>201 προθύμως</td>
<td>0.00002</td>
</tr>
<tr>
<td>202 πρό</td>
<td>0.01106</td>
</tr>
<tr>
<td>203 πρός</td>
<td>0.00005</td>
</tr>
<tr>
<td>204 πρῶς ′ , πρωιθθεν</td>
<td>0.00192</td>
</tr>
<tr>
<td>205 πρῶτως</td>
<td>0.01303</td>
</tr>
<tr>
<td>206 πῶς</td>
<td>0.00152</td>
</tr>
<tr>
<td>207 πώποτε</td>
<td>0.01338</td>
</tr>
<tr>
<td>208 βαμματος</td>
<td>0.00189</td>
</tr>
<tr>
<td>209 σχεδόν</td>
<td>0.00092</td>
</tr>
<tr>
<td>210 σεμνὸς</td>
<td>0.00032</td>
</tr>
<tr>
<td>211 σφόδρα</td>
<td>0.00540</td>
</tr>
<tr>
<td>212 σήμερον</td>
<td>0.00023</td>
</tr>
<tr>
<td>213 σπουδαίως</td>
<td>0.00294</td>
</tr>
<tr>
<td>214 σίν</td>
<td>0.00157</td>
</tr>
<tr>
<td>215 ταχεὺς</td>
<td>0.00522</td>
</tr>
<tr>
<td>216 τάχα</td>
<td>0.00043</td>
</tr>
<tr>
<td>118</td>
<td>καθό</td>
</tr>
<tr>
<td>119</td>
<td>καθότι</td>
</tr>
<tr>
<td>120</td>
<td>καθώς</td>
</tr>
<tr>
<td>121</td>
<td>κατά, κατ’, καθ’</td>
</tr>
<tr>
<td>122</td>
<td>κατέναντι</td>
</tr>
<tr>
<td>123</td>
<td>κάτω</td>
</tr>
<tr>
<td>124</td>
<td>κενώς</td>
</tr>
<tr>
<td>125</td>
<td>κυκλόθεν</td>
</tr>
<tr>
<td>126</td>
<td>λαμπρῶς</td>
</tr>
<tr>
<td>127</td>
<td>λάθρα</td>
</tr>
<tr>
<td>128</td>
<td>λίαν</td>
</tr>
<tr>
<td>129</td>
<td>μάλλον</td>
</tr>
<tr>
<td>130</td>
<td>μακρόθεν</td>
</tr>
<tr>
<td>131</td>
<td>μάλιστα</td>
</tr>
<tr>
<td>132</td>
<td>μεγάλως</td>
</tr>
<tr>
<td>133</td>
<td>μεννόν, μεννύγγε</td>
</tr>
<tr>
<td>134</td>
<td>μετά, μετ’, μεθ’</td>
</tr>
<tr>
<td>135</td>
<td>μεταξύ</td>
</tr>
<tr>
<td>136</td>
<td>μετρίως</td>
</tr>
<tr>
<td>137</td>
<td>μέχρι, μέχρις</td>
</tr>
<tr>
<td>138</td>
<td>μέν</td>
</tr>
<tr>
<td>139</td>
<td>μέντοι</td>
</tr>
<tr>
<td>140</td>
<td>μηδαμῶς</td>
</tr>
</tbody>
</table>
Appendix 11.1


In PRP a_AT0 recent AJO issue NN1 of PRF The_ATO Journal NN1 of PRF Theological_AJO Studies NN2, , Philip NPO Esler NPO paid_VVD me_PNP the_AT0 honour NN1 of PRF writing_VVG a_AT0 lengthy_AJO review_NN1 (_( Esler NPO 1998a_UNC )_) of PRF my_DPS book NN1, , The_AT0 Social_AJO Ethos NN1 of PRF the_AT0 Corinthian_AJO Correspondence_NN1 (_( Horrell NPO 1996 CRD ; ; hereafter_AVO SE NPO )_), , In PRP that_DT0 review_NN1, , rather_CJS21 than_CJS22 focus_VVI on PRF the_AT0 exegetical_AJO sections NN2 of PRF the_AT0 book NN1, , Esler NPO chose_VVD 'to_NN1 concentrate_VVB on PRP the_AT0 theoretical_AJO issues_NN2 ' POS which_DTQ were_VBD raised_VVN in_AVO21 Part_AVO22 1 CRD of PRF the_AT0 study_NN1, , the_AT0 methodological_AJO discussion_NN1 (_( Esler NPO 1998a_UNC : : 255 CRD )_), , Esler NPO's_POS decision_NN1 to_TO0 focus_VVI on PRP these_DT0 issues_NN2 of PRF theory_NN1 and_CJC method_NN1 makes_VVZ a_AT0 response_NN1 of PRF wider_AJC significance_NN1 than_CJS might VM0 otherwise AV0 have_VHI been_VBN the_AT0 case_NN1, , A_AT0 number_NN1 of PRF recent_AJO writers_NN2, , reviewing_VVG the_AT0 state_NN1 of PRF social-scientific_AJO interpretation_NN1 of PRF the_AT0 New_AJO Testament_NN1, , have_VHB highlighted_VVN the_AT0 need_NN1 for PRF discussion_NN1 of PRF the_AT0 methodological_AJO0 and_CJC theoretical_AJO issues_NN2 that_CJT underlie_VVB the_AT0 exegetical_AJO practice_NN1, , Susan NPO Garrett NPO, , for_AVO21 example_AVO22, , insists_VVZ, , 'It_VVN is_VBZ ... ... increasingly AV0 urgent_AJO that_CJT scholars_NN2 of PRF Christian_AJO origins_NN2 engage_VVB in PRP sustained_AJO reflection_NN1 on PRP the_AT0 philosophical_AJO implications_NN2 of PRF the_AT0 perspectives_NN2 and_CJC models_NN2 that_CJT they_PNP choose_VVB to_TO0 employ_VVI ( ( Garrett NPO 1992 CRD : : 93 CRD ) ), , Dale NPO Martin NPO speaks_VVZ of PRF 'a_ITJ spectrum NN1 of PRF opinion_NN1 about PRP what_DTQ precisely AV0 a_AT0 social-scientific_AJO method_NN1 should_VMV0 be_VVI, , with PRP much_DT0 of PRF the_AT0 debate_NN1, , in_PRP his_DPS view_NN1, , centring_VVG 'on_AJO the_AT0 use_NN1 of PRF " " models_NN2 " " derived_VVN from PRP the_AT0 social_AJO sciences_NN2 ' POS ( ( Martin NPO 1993 CRD : : 107 CRD ) ), , Furthermore AV0, , like_PRP Garrett NPO, , Martin NPO maintains_VVZ that_CJT the_AT0 debate_NN1 reflects_VVZ profound_NN1 theoretical_AJO differences_NN2, , particularly AV0 epistemological_AJO ones_NN2 ', POS ( ( Martin NPO 1993 CRD : : 109-10 CRD ) ), ,1.CR0 Discussion NN1 and_CJC reflection_NN1 of PRF the_AT0 social_AJO sciences_NN2 in PRP New_AJO Testament_NN1 interpretation_NN1, , especially AV0 given_VVN the_AT0 sometimes AV0 rather_AVO firm_AJO division_NN1 between PRP two_CRD groups_NN2 who_PNP have_VHB come_VVN to_TO0 be_VBI referred_VVN to_TO0 as_AVO 'the_VVI social_AJO historians_NN2 ' POS and_CJC the_VVB social_AJO scientists_NN2 ', POS, , In PRP the_AT0 following_AJO essay_NN1, , I_PNP shall_VM0 attempt_VVI to_TO0 respond_VVI to_PRP Esler NPO's_POS arguments_NN2 concerning_PRP the_AT0 use_NN1 of PRF models_NN2, , the_AT0 validity_NN1 of PRF Berger NPO and_CJC Luckmann' NPO zz0 theory_NN1 of PRF the_AT0 social_AJO construction_NN1 of PRF reality_NN1, , and_CJC, , finally AV0 and_CJC more_AV0 briefly_AV0, , the_AT0 major_AJO argument_NN1 of PRF SE NPO 1 CRD, , The_AT0 Use_NN1 of PRF Models_NN2, ,

255
In PRP SENPO I PNP took_VVD issue_NNl with_PRP a_ATO t i & f i  AVO representedVVN view N N l H _ C J T  a_ATO social-scientificAJO approach NNl to PRP the ATO New_AJO Testament NNl essentially_AVV involves_VVZ the_AT0 use NNl of PRF 'models_NN2 ' POS ( ( SE_NP0 9-18 CRD ) ) _ . _ I CRD argued-and_CJC will_VM0 not_XX0 tin ip f AVO repeat_VVI those_DT0 arguments_NN2 here-that_CJT there_EX0 were_VBD a_AT0 number_NN1 of PRF problems_NN2 with_PRP such_DT0 a_AT0 conception_NN1 of PRF social-scientific_AJO method_NN1 and_CJC maintained_VVD that_CJT it_PNP was_VBD certainly_AVV not_XX0 the_AT0 only_AJO way_NN1 to_TO0 approach_VVI the_AT0 New_AJO Testament_NN1 'sociologically'_NN1 _ . _ Not_XX0 the_AT0 least_DT0 of PRF these_DT0 problems_NN2 _ , _ it_PNP seems_VVZ to_PRP me_PNP _ _ is_VBZ the_AT0 way_NN1 in PRP which_DTQ a_AT0 model-based_AJO approach_NN1 can_VM0 lead_VVI to PRP historically_AVV and_CJC culturally_AVV variable_AJO evidence_NN1 being_VBG interpreted_VVN through_PRP the_AT0 lens_NN1 of PRF a_AT0 generalized_AJO model_NN1 of PRF social_AJO behaviour_NN1 _ . _ The_AT0 rich_AJO diversity_NN1 of PRF human_AJO behaviour_NN1 is_TBZ thereby_AVV homogenized_VVN and_CJC explained_VVN in PRP31 terms_PRP32 of_PRP33 what_DTQ is_VBZ 'typical' which_DTQ can VM0 actually_AVV be_VBI no_AT0 real_AJO explanation_NN1 of AVN21 AVN22 ( _ ( cf._VVB SE_NN1 _ _ 287-89 CRD ) ) _ _ . _ Although_CJS Esler_NP0 is_VBZ conscious_AJO of_PRF these_DT0 dangers_NN2 _ _ and_CJC insists_VVZ on PRP the_AT0 need_NN1 for PRP sensitivity_NN1 to PRP cultural_AJO variation_NN1 and_CJC specificity_NN1 ( _ ( see_VVB Esler_NP0 1998b_UNC : 4-5 CRD ) ) _ _ _ his_DPS most_AVV recent_AJO work_NN1 still_AVV exhibits_VVZ the_AT0 problem_NN1 _ _ as_CJS will_VM0 be_VBI shown_VVN below_NN0

Explanation of Tag Symbols

UCREL CLAWS5 Tagset

!  PUN
"  PUQ
(  PUL
)  PUR
,  PUN
-  PUN
.  PUN
...  PUN
;  PUN
?  PUN

AJ0  adjective (unmarked) (e.g. GOOD, OLD)
AJC  comparative adjective (e.g. BETTER, OLDER)
AJS  superlative adjective (e.g. BEST, OLDEST)
AT0  article (e.g. THE, A, AN)
AV0  adverb (unmarked) (e.g. OFTEN, WELL, LONGER, FURTHEST)
AVP  adverb particle (e.g. UP, OFF, OUT)
AVQ  wh-adverb (e.g. WHEN, HOW, WHY)
CJC  coordinating conjunction (e.g. AND, OR)
CJS  subordinating conjunction (e.g. ALTHOUGH, WHEN)
CJT  the conjunction THAT
<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRD</td>
<td>cardinal numeral (e.g. 3, FIFTY-FIVE, 6609) (excl ONE)</td>
</tr>
<tr>
<td>DPS</td>
<td>possessive determiner form (e.g. YOUR, THEIR)</td>
</tr>
<tr>
<td>DT0</td>
<td>general determiner (e.g. THESE, SOME)</td>
</tr>
<tr>
<td>DTQ</td>
<td>wh-determiner (e.g. WHOSE, WHICH)</td>
</tr>
<tr>
<td>EX0</td>
<td>existential THERE</td>
</tr>
<tr>
<td>ITJ</td>
<td>interjection or other isolate (e.g. OH, YES, MHM)</td>
</tr>
<tr>
<td>NN0</td>
<td>noun (neutral for number) (e.g. AIRCRAFT, DATA)</td>
</tr>
<tr>
<td>NN1</td>
<td>singular noun (e.g. PENCIL, GOOSE)</td>
</tr>
<tr>
<td>NN2</td>
<td>plural noun (e.g. PENCILS, GEESE)</td>
</tr>
<tr>
<td>NP0</td>
<td>proper noun (e.g. LONDON, MICHAEL, MARS)</td>
</tr>
<tr>
<td>NULL</td>
<td>the null tag (for items not to be tagged)</td>
</tr>
<tr>
<td>ORD</td>
<td>ordinal (e.g. SIXTH, 77TH, LAST)</td>
</tr>
<tr>
<td>PNI</td>
<td>indefinite pronoun (e.g. NONE, EVERYTHING)</td>
</tr>
<tr>
<td>PNP</td>
<td>personal pronoun (e.g. YOU, THEM, OURS)</td>
</tr>
<tr>
<td>PNQ</td>
<td>wh-pronoun (e.g. WHO, WHOEVER)</td>
</tr>
<tr>
<td>PNX</td>
<td>reflexive pronoun (e.g. ITSELF, OURSELVES)</td>
</tr>
<tr>
<td>POS</td>
<td>the possessive (or genitive morpheme) 'S or '</td>
</tr>
<tr>
<td>PRF</td>
<td>the preposition OF</td>
</tr>
<tr>
<td>PRP</td>
<td>preposition (except for OF) (e.g. FOR, ABOVE, TO)</td>
</tr>
<tr>
<td>PUL</td>
<td>punctuation - left bracket (i.e. ( or [ )</td>
</tr>
<tr>
<td>PUN</td>
<td>punctuation - general mark (i.e. .!,:;-? ...)</td>
</tr>
<tr>
<td>PUQ</td>
<td>punctuation - quotation mark (i.e. ' 1  &quot;)</td>
</tr>
<tr>
<td>PUR</td>
<td>punctuation - right bracket (i.e. ) or ] )</td>
</tr>
<tr>
<td>TOL</td>
<td>infinitive marker TO</td>
</tr>
<tr>
<td>UNO</td>
<td>&quot;unclassified&quot; items which are not words of the English lexicon</td>
</tr>
<tr>
<td>VBB</td>
<td>the &quot;base forms&quot; of the verb &quot;BE&quot; (except the infinitive), i.e. AM, ARE</td>
</tr>
<tr>
<td>VBD</td>
<td>past form of the verb &quot;BE&quot;, i.e. WAS, WERE</td>
</tr>
<tr>
<td>VBG</td>
<td>-ing form of the verb &quot;BE&quot;, i.e. BEING</td>
</tr>
<tr>
<td>VBI</td>
<td>infinitive of the verb &quot;BE&quot;</td>
</tr>
<tr>
<td>VBN</td>
<td>past participle of the verb &quot;BE&quot;, i.e. BEEN</td>
</tr>
<tr>
<td>VBZ</td>
<td>-s form of the verb &quot;BE&quot;, i.e. IS, 'S</td>
</tr>
<tr>
<td>VDB</td>
<td>base form of the verb &quot;DO&quot; (except the infinitive), i.e.</td>
</tr>
<tr>
<td>VDD</td>
<td>past form of the verb &quot;DO&quot;, i.e. DID</td>
</tr>
<tr>
<td>VDG</td>
<td>-ing form of the verb &quot;DO&quot;, i.e. DOING</td>
</tr>
<tr>
<td>VDI</td>
<td>infinitive of the verb &quot;DO&quot;</td>
</tr>
<tr>
<td>VDN</td>
<td>past participle of the verb &quot;DO&quot;, i.e. DONE</td>
</tr>
<tr>
<td>VDZ</td>
<td>-s form of the verb &quot;DO&quot;, i.e. DOES</td>
</tr>
<tr>
<td>VHB</td>
<td>base form of the verb &quot;HAVE&quot; (except the infinitive), i.e. HAVE</td>
</tr>
<tr>
<td>VHD</td>
<td>past tense form of the verb &quot;HAVE&quot;, i.e. HAD, 'D</td>
</tr>
<tr>
<td>VHG</td>
<td>-ing form of the verb &quot;HAVE&quot;, i.e. HAVING</td>
</tr>
<tr>
<td>VHI</td>
<td>infinitive of the verb &quot;HAVE&quot;</td>
</tr>
<tr>
<td>VHN</td>
<td>past participle of the verb &quot;HAVE&quot;, i.e. HAD</td>
</tr>
<tr>
<td>VHZ</td>
<td>-s form of the verb &quot;HAVE&quot;, i.e. HAS, 'S</td>
</tr>
<tr>
<td>VM0</td>
<td>modal auxiliary verb (e.g. CAN, COULD, WILL, 'LL)</td>
</tr>
<tr>
<td>VVB</td>
<td>base form of lexical verb (except the infinitive)(e.g. TAKE, LIVE)</td>
</tr>
<tr>
<td>VVD</td>
<td>past tense form of lexical verb (e.g. TOOK, LIVED)</td>
</tr>
<tr>
<td>VVG</td>
<td>-ing form of lexical verb (e.g. TAKING, LIVING)</td>
</tr>
<tr>
<td>VVI</td>
<td>infinitive of lexical verb</td>
</tr>
<tr>
<td>VVN</td>
<td>past participle form of lex. verb (e.g. TAKEN, LIVED)</td>
</tr>
<tr>
<td>VVZ</td>
<td>-s form of lexical verb (e.g. TAKES, LIVES)</td>
</tr>
<tr>
<td>XX0</td>
<td>the negative NOT or N'T</td>
</tr>
<tr>
<td>ZZ0</td>
<td>alphabetical symbol (e.g. A, B, c, d)</td>
</tr>
</tbody>
</table>
BIBLIOGRAPHY

Computer Packages


Packages used:

cluster. version (1.11.0) Martin Maechler, based on S original by Peter Rousseeuw, Anja.Struyf@uii.ua.ac.be and Mia.Hubert@uii.ua.ac.be, and initial R port by Kurt.Hornik@R-project.org.

depend. version (1.0-3) Maria L. Rizzo and Gabor J. Szekely.


randomForest. version (4.5-16) Fortran original by Leo Breiman and Adele Cutler, R port by Andy Liaw and Matthew Wiener.

rpart. version (3.1-29) Terry M Therneau and Beth Atkinson. R port by Brian Ripley.

varSelRF. Version (0.6-2) Ramon Diaz-Uriarte.

Corel Paradox 9 Relational Database Software. Corel Corporation.
OVW available from makarenv@ere.umontreal.ca.

Literature


Biber, D. "Historical Patterns for the Grammatical Marking of Stance". Journal of Historical
Brenton, L.C.L. The Septuagint with Apocrypha: Greek and English. Zondervan: Grand Rapids 1992
Campenhausen von, H. Ecclesiastical Authority and Spiritual Power in the Church of the First Three Centuries. Hendrilson: Peabody 1997
Chaski, C.E. "Empirical Evaluations of Language-Based Author Identification Techniques". Forensic Linguistics 8 (2001)

259
Dibelius, M.  *Der Hirt des Hermas*. in Die Apostolischen Väter Mohr Siebeck: Tübingen 1923
Fiorenza, E.S.  "Rhetorical Situation and Historical Reconstruction in 1 Corinthians"  *NTS* 33 (1987) pp.386-403
Friedman, J.H. "On Bias, Variance, 0/1 - Loss, and the Curse-of-Dimensionality". Data Mining and Knowledge Discovery 1 (1997) pp.55-77

Gager, J. Kingdom and Community: The Social World of Early Christianity. 1975


Gieschen, C.A. Angelomorphic Christology. Brill: Leiden


Harris, J.R. Hermas in Arcadia and other Essays. Cambridge University Press: Cambridge 1896


261


Maier, H.O. The Social Setting of the Ministry as Reflected in the Writings of Hermas, Clement and Ignatius. Wilfrid Laurier University Press: Canada 1991

Mommsen, T. *The History of Rome: The Provinces From Caesar to Diocletian*. (2 vols): London 1886
Neusner, J. (ed) *Christianity, Judaism and other Greco-Roman Cults*.: Leiden 1975


Robinson, J.M. *Trajectories Through Early Christianity*. 1979


Sherwin-White, A.N. "The Early Persecutions and the Roman Law Again". *JTS NS* 3 (1952) pp.199-213


Streeter, B.H. *The Primitive Church*. London 1923

Sundberg, A.C.  "Canon Muratori: A Fourth Century List".  *HTR* 66 (1973) pp.1ff
Whittaker, M. *der Hirt des Hermas*. Akademie-Verlag: Berlin 1956
Wilson, W.J. "The Career of the Prophet Hermas".  *HTR* 20 (1927) pp.21-62