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

UMI Number: U584830

UMI

Dissertation Publishing
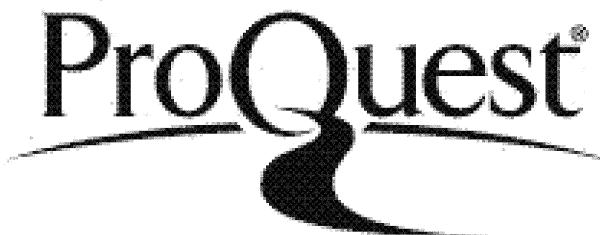
ProQuest

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

DECLARATION

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed .... (candidate)

Date ....

STATEMENT 1

This thesis is the result of my own investigations, except where otherwise stated. Where correction services have been used, the extent and nature of the correction is clearly marked in a footnote(s).

Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed .... (candidate)

Date ....

STATEMENT 2

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed .... (candidate)

Date ....

NB: *Candidates on whose behalf a bar on access has been approved by the University (see Note 9), should use the following version of Statement 2:*

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loans **after expiry of a bar on access approved by the University of Wales on the special recommendation of the Member Institution concerned.**

Signed .... (candidate)

Date ....

ACKNOWLEDGMENTS

ABBREVIATIONS

ANRW   *Aufstieg und Niedergang der römischen Welt.* Edited by Hildegard Temporini and Wolfgang Haase

ATR   *Anglican Theological Review*

Bauer   *A Greek-English Lexicon Of The New Testament (1958) translated by W.Arnt and W.Gingrich*

BJRL   *Bulletin of the John Rylands University Library of Manchester*

CBQ   *Catholic Biblical Quarterly*

CH   *Computers and the Humanities*

CIG   *Corpus Inscriptionum Graecarum*

EH   *Ecclesiastical History* (Eusebius)

HTR   *Harvard Theological Review*

JBL   *Journal of Biblical Literature*

JECS   *Journal of Early Christian Studies*

JEH   *Journal of Ecclesiastical History*

JHP   *Journal of Historical Pragmatics*

JML   *Journal of Machine Learning*

JNTS   *Journal of New Testament Studies*

JSNT   *Journal for the Study of the New Testament ITS Journal of Theological Studies*

JTS NS   *Journal of Theological Studies, New Series*

LSJ   *Liddell, Scott and Jones Greek-English Lexicon, (1940)*

NovT   *Novum Testamentum*

NTS   *New Testament Studies*

SA   *Sociological Analysis*

TDNT   *Theological Dictionary of the New Testament* translated by Kittel-Friedrich

VC   *Vigiliae Christianae*

ZNW   *Zeitschrift für die neutestarnentlich Wissenschaft und die Kunde der ällteren Kirche*

ZTK   *Zeitschrift für Theologie und Kirche*

ZWT   *Zeitschrift für Wissenschaft Theologie*

CONTENTS

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[1]. 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[2], 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[3], 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.*

1

(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.

---

4   Richardson, E.C. (tr.) *Nicene and Post-Nicene Fathers.* 3, pp.359-384
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 (א), 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[5]. 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 א, and the text from A were supplemented by the Vulgate, from which Bishop Lightfoot published his translation in 1890[6]. 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 א (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[7].

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

5   Turner, C.H. "The Shepherd of Hermas and the Problem of its Text." *JTS* (1920) 21, pp.193-209
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 (ℵ) 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 ℵ is thought to be the work of at least three scribes and shows the corrections of several editors, Kirsopp Lake does not consider ℵ 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."*[9]

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."*[10] (bold text mine)

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[11], Johannine[12] and Petrine[13] 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[14] (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."*[15]

---

9   Streeter, B.H. *The Primitive Church.* p.203
10   Elliott, J. *Social-Scientific Criticism of the New Testament: An Introduction.* p.9
11   Meeks, W.A. *The First Urban Christians: The Social World of the Apostle Paul.*
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.*
14   Robinson, J.M.& Koester, H. (eds.) *Trajectories Through Early Christianity.*
15   Oakes, P. (ed.) *Rome in the Bible and the Early Church.* P.xvii

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.
17 Brown, R.E.&Meier, J.P. *Antioch and Rome.* and Gregory, A. "Disturbing Trajectories: 1 Clement, the Shepherd of Hermas and the Development of Early Roman Christianity" in Oakes, P.(ed.) *Rome in the Bible and the Early Church.* pp.142-166
18 Simmel, G. *Conflict: The Web of Group-Affiliations.* pp16ff

6

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[19] 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[20] who was able to outline a scientific, practical methodology for a statistical-linguistic study of Greek texts. A significant portion of this thesis is

19 Coleborne, W. "The Shepherd of Hermas: A Case for Multiple Authorship and Some Implications." *Studia Patristica 10* (1970) 107 pp.65-70
20 Mealand, D.L. "Style, Genre and Authorship in Acts, The Septuagint and Hellenistic Historians." *Journal of Association for Literary and Linguistic Computing* (1999) 14.4 pp.479-506

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 (῎Ορασις α) is concerned mainly with a vision Hermas receives of an elderly woman *(γυνή πρεσβῦτις).* She sits on a great white chair *(καθέδραν λευκν μεγάλην)* and reads to Hermas from a book that she has brought with her. After completing the reading she departs towards the East *(πρός τήν ἀνατολήν).*

Vision 2 (῎Ορασις 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 ὑπό νεανίσκος),* given to Hermas while he slept. He reveals to Hermas that the elderly woman is 'The Church'.

Vision 3 (῎Ορασις γ) recounts an appearance of the elderly woman once more, and she and Hermas sit together on an ivory couch covered with linen *(συμψέλιον ἐλεφάντινον...καί ἐπάνω λέντιον 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 *(περί τῶν τριῶν μορφῶν, ἐν αἷς μοι ἐνεφανίσθη).* The young man supplies the information.

Vision 4 (῎Ορασις 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 (᾽Απокάλυψις 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 5th 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.

ὤφθη δέ μοι ἀδελφοί τῇ μέν πρώτῃ ὁράσει
τῇ περυσινῇ λίαν πρεσβυτέρα καί ἐν καθέδρᾳ
καθημένη. τῇ δέ ἑτέρα ὁράσει τήν μέν ὄψιν
νεωτέραν εἶχεν τήν δέ σάρκα καί τάς τρίχας
πρεσβυτέρας καί ἑστηκυῖα μοι ἐλάλει ἱλαρωτέρα
δέ ἦν ἤ τό πρότερον. τῇ δέ τρίτῃ ὁράσει ὄλη
νεωτέρα καί κάλλει ἐκπρεπεστάτη μόνας δέ τάς
τρίχας πρεσβυτέρας εἶχεν ἱλαρά δέ εἰς τέλος
ἦν καί ἐπί συμψελίου καθημένη. (V3.10.3-5)

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 (4th C.), but Codex Athous (15th C.) heads Vision 5 as ῾Όρασις ε, probably an editorial harmonization. However, ᾽Απολάλυψις ε is probably the best reading.
22 Carlini, A. *Papyrus Bodmer XXXVIII. Erma: Il Pastore[Ia – IIIa visione]*[Bibliotheca Bodmeriana; Cologny-Genève: Foundation Martin Bodmer, 1991]. Quoted in Osiel, C. *The Shepherd of Hermas* p.1. The break is mid V3.13.4.
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**.
25 Kirkland, A. "The Literary History of The Shepherd of Hermas." *The Second Century* (1992) 9 pp.87-102

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. Osiek[26] 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 (ἀποβαλεῖς πᾶσαν πονηρίαν) in order to be clothed with every righteous virtue (ἐνδύσῃ πᾶσαν ἀρετήν 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 *(ἁπλότης)* and innocence *(ἀκακία)*. This *Mandate* also contains a number of verbal parallels with the *Didache* eg.

Πᾶσιν γάρ ὁ θεός δίδοσθαι θέλει ἐκ τῶν ἰδίων δωρημάτων. (M2.1.4)

Πᾶσι γάρ 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 *(ἀλήθεια)*.

Mandate 4 seems to have a dislocation of the text. M4.1.1-11 deals with questions about adultery, while M4.4.1-4 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 *(ἱλαρότης)* and well-being *(εὐθηνία)*, 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 *(ίλαρότης)*, 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 *(έπιθυμία)* before concluding with M12.3.2. However, Hennas 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*[28].

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 τοῦτο γάρ ἐπλούτισεν ὑμᾶς ὁ δεσπότης, ἵνα
> ταύτας τάς διακονίας 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 καρποί φανεροί)*.

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[29] 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**.*[30] (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 Osiek[31], "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):

$$\dot{\alpha}\lambda\lambda\dot{\alpha} \ \kappa\alpha\dot{\iota} \ \dot{\alpha}\pi o\delta\acute{\epsilon}\chi\epsilon\tau\alpha\iota \ \tau\grave{\eta}\nu \ \tau o\hat{\upsilon} \ \Pi o\iota\mu\acute{\epsilon}\nu o\varsigma \ \gamma\rho\alpha\phi\acute{\eta}\nu$$

$$\lambda\acute{\epsilon}\gamma\omega\nu, \ \kappa\alpha\lambda\hat{\omega}\varsigma \ o\hat{\upsilon}\nu \ \dot{\eta} \ \gamma\rho\alpha\phi\grave{\eta} \ \dot{\eta} \ \lambda\acute{\epsilon}\gamma o\upsilon\sigma\alpha \ \pi\rho\hat{\omega}\tau o\nu$$

$$\pi\acute{\alpha}\nu\tau\omega\nu \ \pi\acute{\iota}\sigma\tau\epsilon\upsilon\sigma o\nu \ \acute{o}\tau\iota \ \epsilon\hat{\iota}\varsigma \ \acute{\epsilon}\sigma\tau\iota\nu \ \dot{o} \ \theta\epsilon\grave{o}\varsigma \ \dot{o} \ \tau\grave{\alpha}$$

$$\pi\acute{\alpha}\nu\tau\alpha \ \kappa\tau\acute{\iota}\sigma\alpha\varsigma \ \kappa\alpha\dot{\iota} \ \kappa\alpha\tau\alpha\rho\tau\acute{\iota}\sigma\alpha\varsigma^{32}$$

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*[34]:

> *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."[35]

4.   Tertullian (c.160 – c.225 AD)

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

---

34   Origen *De Principiis* IV:1:11-12 in Roberts, A. and Donaldson,J. (eds.) *Ante-Nicene Fathers*. IV, p.359-360
35   Quoted in Wilson, J.C. *Five problems in the Interpretation of the Shepherd of Hermas*. p.63

authoritative. 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...*[36]

He dislikes the "shepherd of adulterers" because it can countenance and advocate repentance even after divorce for adultery[37]. 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 ἐπὶ τέλει
> προσρήσεσιν τῆς πρὸς ῾Ρωμαίους μνήμην
> πεποίηται μετὰ τῶν ἄλλων καὶ ῾Ερμᾶ, οὖ φασιν
> ὑπάρχειν τὸ τοῦ Ποιμένος βιβλίον, ἱστέρον ὡς καὶ
> τοῦτο πρὸς μέν τινων ἀντιλέλεκται, δι᾽ οὓς οὐκ ἂν
> ἐν ὁμολογουμένοις τεθείη, ὑφ᾽ ἑτέρων δὲ
> ἀναγκαιότατον οἷς μάλιστα δεῖ στοιχειώσεως
> εἰσαγωγικῆς, κέκριται· ὅθεν ἤδη καὶ ἐν
> ἐκκλησίαις ἴσμεν αὐτὸ δεδημοσιευμένον, καὶ τῶν
> παλαιτάτων δὲ συγγραφέων κεχρημένους τινὰς
> αὐτῷ κατείληφα.[38]

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.

---

36   Tertullian, *De Pudicitia.* XX verse 2 in Roberts, A. and Donaldson,J. (eds.) , *The Ante-Nicene Fathers*, IV, p.168
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[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[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.
40   Robbins, G.A. "Muratorian Fragment." In *Anchor Bible Dictionary.* Freedman, D.N. (ed.) 4 p.929. Also, Koch, H. "Zu A. v. Harnacks Beweis für den amtlichen römischen Ursprung des Muratorischen Fragments." *ZNW* (1926) 25, pp.154-160. Quoted in Sundberg, A. C. "Canon Muratori: A Fourth Century List", *HTR* (1973) 66 pp.1-41

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[41]. This has implications with regard to its dating. Hill[42] deals with the Latin word 'catafrygum' (line 84 of the Fragment) and Hahneman[43] argues that this is a translation of the Greek καταφρυγας 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[44] 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[45] 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
42 Hill, C.E. "The Debate Over The Muratorian Fragment and the Develepoment of the Canon." *Westminster Theological Journal* (1995) 57, pp.473ff
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.
45 Zahn, "Muratorian Canon." *The New Schaff-Herzog Religious Encyclopaedia*. p.54 . Quoted in Hahneman.

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[46] 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

46 Ferguson,E "Review of Geoffrey Mark Hahneman's The Muratorian Fragment and the Development of the Canon." *JTS* (1993) 44, p.692

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. Eusebius[48] may allude to this in a short comment that he makes:

> τῶν δ' ἀμφὶ τὸν Μοντανὸν καί 'Αλκιβιάδην καὶ
> Θεόδοτον περί τὴν Φρυγίαν ... ἃς ἐν δεσμοῖς ἔτι
> ὑπάρχοντες τοῖς... 'Ελευθέρῳ τῷ τότε 'Ρωμαίων
> ἐπισκόπῳ, τῆς τῶν ἐκκλησιῶν εἰρήνης ἕνεκα
> πρεσβεύοντες.

---

47 Athanasius *Ep. Fest. 39* (367AD) in Schaff,P. (ed.) *Nicene and Post Nicene Fathers.* IV, p.964
48 Lake,K. *Eusebius Ecclesiastical History.* 5.3.4. p.443

If "the most plausible interpretation[49]" 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.[50]*

(Attributed to Priscilla or Quintilla[51]).

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[52] 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 Epiphanus.* 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[53], for example, believes Montanism's effect on canon creation to be over-estimated. 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:

Τὴν δὲ καθόλου καὶ πᾶσιν τὴν ὑπὸ τὸν οὐρανὸν
ἐκκλησίαν βλασφημεῖν διδάσκοντος τοῦ
ἀπηνθαδισμένου πνεύματος, ὅτι μήτε τιμὴν μήτε
πάραδον εἰς αὐτὴν τὸ ψευδοπροφητικὸν ἐλάμβανε
πνεῦμα , τῶν γὰρ κατὰ τὴν 'Ασίαν πιστῶν
πολλάκις καὶ πολλαχῇ τῆς 'Ασίας εἰς τοῦτο
συνελθόντων καὶ τοὺς προσφάτους λόγους

53   Paulsen, H. "Die Bedeutung des Montanismus für die Herausbildung des kanons." *Vigiliae Christianae* (1978) 32, pp.19-52

ἐξετασάντων καὶ βεβήλους ἀποφηνάντων καὶ
ἀποδοκιμασάντων τὴν αἵρεσιν, οὕτω δὴ τῆς τε
ἐκκλησίας ἐξεώσθησαν καὶ τῆς κοινωνίας
εἴρχθησαν. [54]

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[55].

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[56] 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[57] 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
56   Sundberg,A.C. "Canon Muratori: A Fourth Century List". HTR (1973) 66, p.1-41.
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 commentators58. 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[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[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[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
61   Coleborne, W. "A linguistic Approach to the Problem of Structure and Composition of the Shepherd of Hermas". *Colloquium* (1969) 3, pp.133-142.

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 failure62. 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.

Chapter 2

The Quest For Authorial Style

And

The Rise Of Statistical Linguistics.

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[63]. 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:

---

63   Binongo, J.N.G. "Who Wrote The 15th Book of Oz? An Application of Multivariate Analysis to Authorship Attribution." *Chance* (2003) 16 pp.19-17.

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[67], for example, argues that by only taking the "*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 "*the quantification of the dimension of style*". This is an important point that cannot be stressed enough, **quantification methods do not provide the final, or only, solution in an analysis; context and common sense are equally as important.**

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[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[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

---

67   Botha, J.E. "Style in the New Testament: The need for Serious Reconsideration." *JSNT* (1991) 43 pp.71-87
68   Biber,D. *Variations Across Speech and Writing.*
69   Chaski, C.E. "Empirical Evaluations of Language-Based Author Identification Techniques." *Forensic Linquistics,* (2001) pp.8ff

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. Grisso[70] 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 marker[71]. 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.
71   Morton, A. Q. "The Authorship of Greek Prose." *Journal of the Royal Statistical Society* (1965) 128 pp.169 – 233

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:

	Galatians	John
	75	16
	33	15
	16	4
	15	13
	7	12
	4	8
	9	6
	15	11
	14	39
	39	22
	54	11
	28	15
	11	16
	10	11
	12	10
	22	8
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.

Density Curves Of Small Sample Size of Sentence Lengths From Two New Testament Texts

N = 16   Bandwidth = 5.809
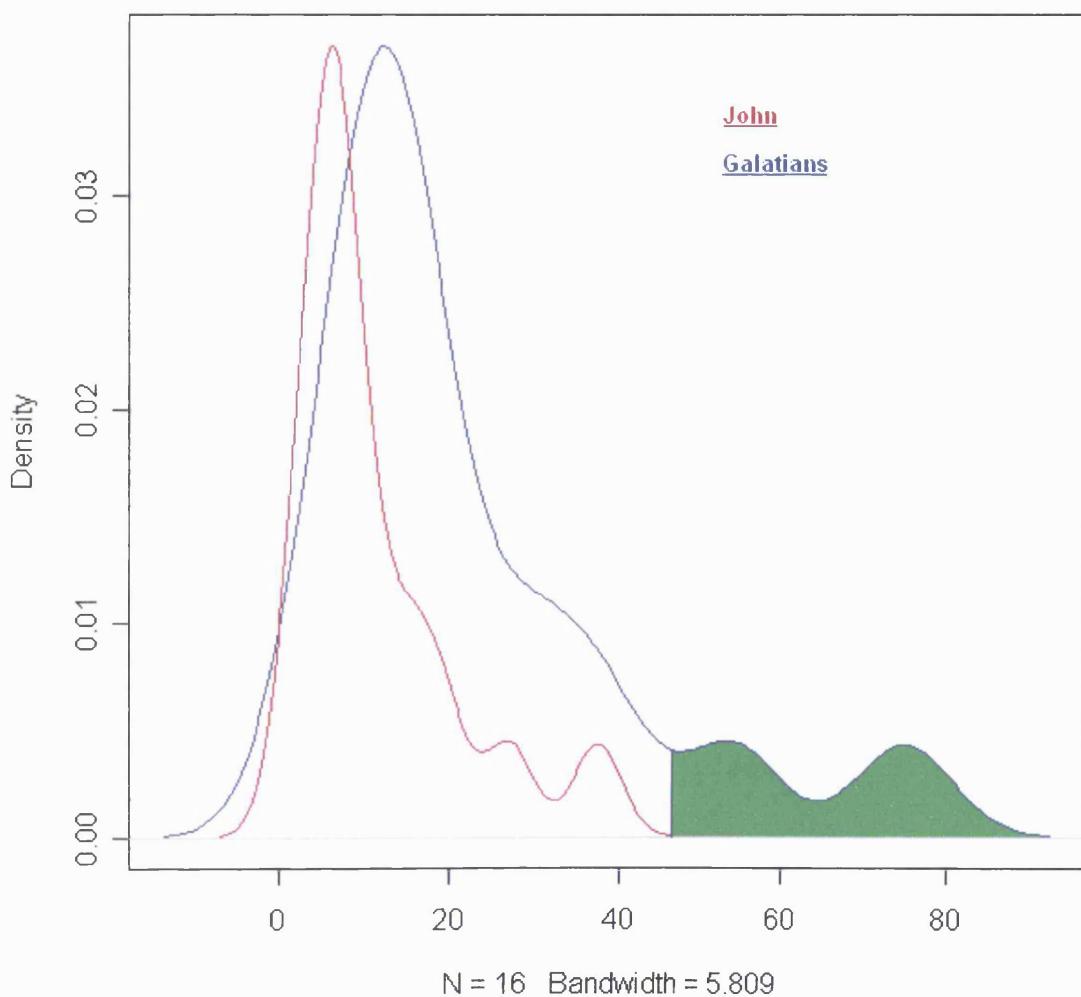
This diagram shows that sentences in the green shaded area are the only ones that

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[72]. 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[73] 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:

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

Where i represents author.

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

Text	TTR	N
Galatians	0.3021	2230
John	0.1778	15641
Romans	0.284	7109
Genesis	0.1475	32553
Acts	0.2858	18379
Revelation	0.2175	9841
Exodus	0.167	24818
Mark	0.2602	11313
1 Corinthians	0.291	6829
Jeremiah	0.1839	28946

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.

---

73   Good, I.J. "The Population Frequencies of Species and the Estimation of Population Parameters." Biometrika (1953) 40 pp.237-264.

Type Token Ratio as a Function of Text Length
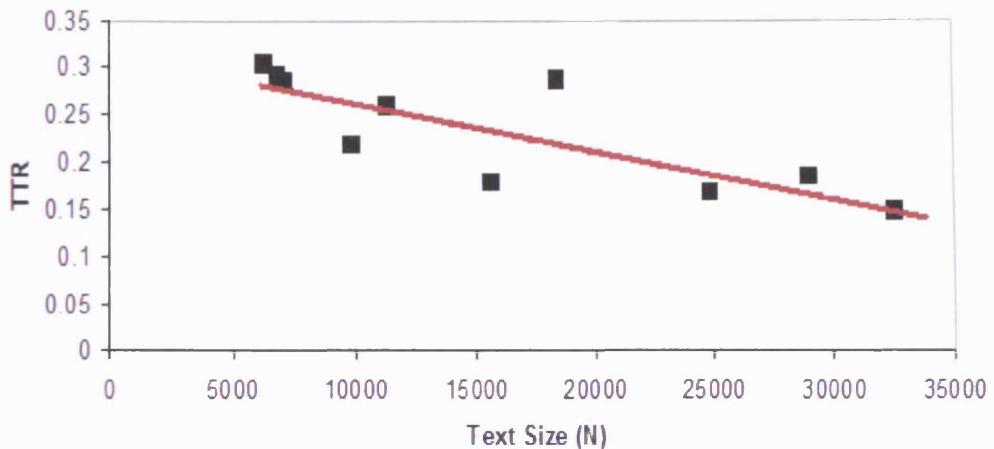
Fig.2.2 Plot of type token ratio against text length. It becomes obvious that there is a relationship between text length and TTR shown by the linear regression line (in red).

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)[74].

Guiraud[75] modified the TTR using the square root of text size:

$$R = \frac{V(N)}{\sqrt{N}}$$

Herdan[76] proposed a simple power function of text size:

---

74   Tweedie F.J. "How Variable May a Constant be? Measures of Lexical Richness in Perspective." *Computers and the Humanities* (1998) 32 pp.323-352.

75   Guiraud H. *Les Caracteres Statistiques du Vocabulaire.*

76   Herdan G. *Quantitative Linguistics.*

$$V(N) = N^c$$

where
$$C = \frac{\log\left(V\left(N\right)\right)}{\log\left(N\right)}$$

Dugast[77] suggested a simple power function of log of text size:

$$V(N) = \log^k (N)$$

where
$$k = \frac{\log\left(V(N)\right)}{\log(\log(N))}$$

Bruner[78] changed the base of the simple power function of the log of a power of the text size!

$$V(N) = \log_N^a (W)$$

where
$$W = N^{v(N)^{-a}}$$

and a 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[79].

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

77   Dugast D. *Vocabulaire et Stylistique.*
77   BrunetE. *Vocabulaire de Jean Giraudoux.*

79   Prusscha,H. "Statistical Models for Vocabulary and Text Length with an Application to the N.T. Corpus" *Literary and Linguistic Computing* (1998), 4, pp.195-198

markers able to distinguish authorship. A radical change in thinking was required. Burrows[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

---

80   Burrows J.F. "'An Ocean Where Each Kind...': Statistical Analysis and Some Major Determinants of literary Style." *CH* (1989) 23 pp.309-321

'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.. [81]

Radford gives a simple test to distinguish between contentives and functors. He suggested that if a word has an antonym it is a contentive[82]. 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

---

81 Gelderen, Elly van. *The Rise of Functional Categories.*
82 Radford, A. *Syntax. p38*

see seven women round the tower?" is a question since there is also a question earlier in the passage. The noun 'tower' is used twice in the passage, firstly in the question, then it undergoes a transformation so that it becomes the subject of the clause, "This tower is being supported by them". The fore grounding of the noun requires the use of the passive verbal form 'is being supported' and the subjects of the original statement (the seven women) are placed into a prepositional phrase and reduced to a pronoun, "by them". Once the tower has achieved importance in the mind of the reader/writer it can be seen how other elements in the passage reinforce its importance. If we now colour code the words that are repeated, we can see two important lines of development.

> *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:

> *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 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, ie. 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[83] 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

---

83   Longobardi, G. "Reference and Proper Names." *Linguistic Inquiry* (1994) 25 pp.609-666

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."*[84] 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[85] 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

---

84   Radford, A. *Syntax.* p.267
85   Baayen, R.H. Tweedie, F.J. & van Halteren, H. "Outside the Cave of Shadows: Using Syntactic Annotation to Enhance Authorship Attribution." *Literary and Linguistic Computing.* (1969) 11 pp. 121-131

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[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

Chapter 3

Model Building: Specification And Data Preliminaries.

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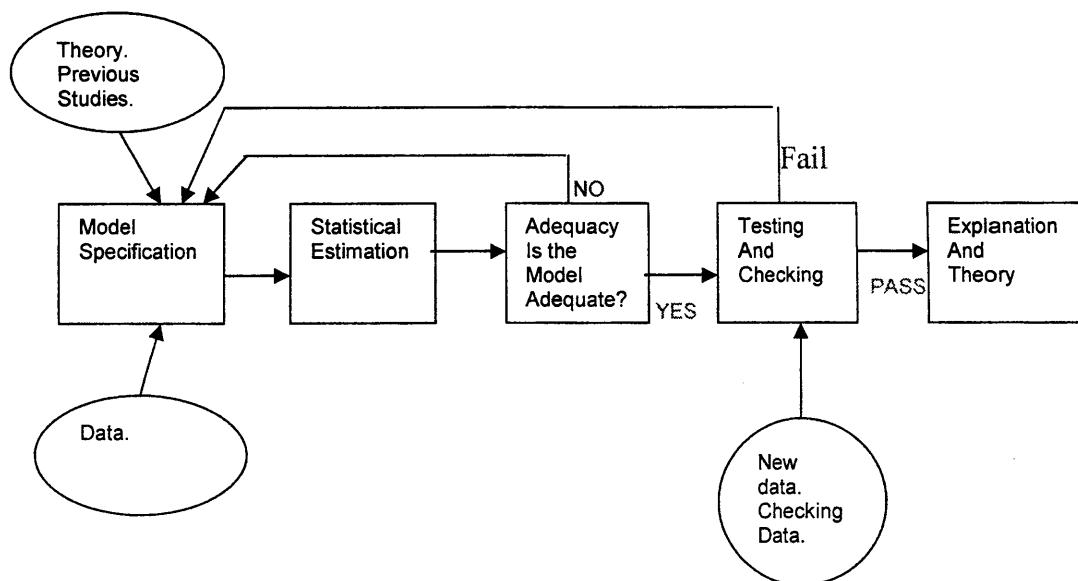
This defines a procedure that can be followed for all model-building problems. In our case, the problem can be defined as:

"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[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 ie. 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:

"In a choice among competing hypotheses, other things being equal, the simplest is preferable."

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.

Data Collection And Manipulation Before Analysis

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:

- **GENESIS.** This text was chosen because we can claim with absolute

---

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 material[88] 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

---

88   Van Den Hoek,A. "Techniques of Quotation in Clement Of Alexandria: A review of Ancient Literary Working" *VC* (1996) 50 pp.223-243

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 καί ('and') has a mean count of 230 per 3000 word block. On the other hand, words such as κατέναντι ('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 κατέναντι 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. $9e^-36$. 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 mRs 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 $\delta\iota\alpha$ ('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.
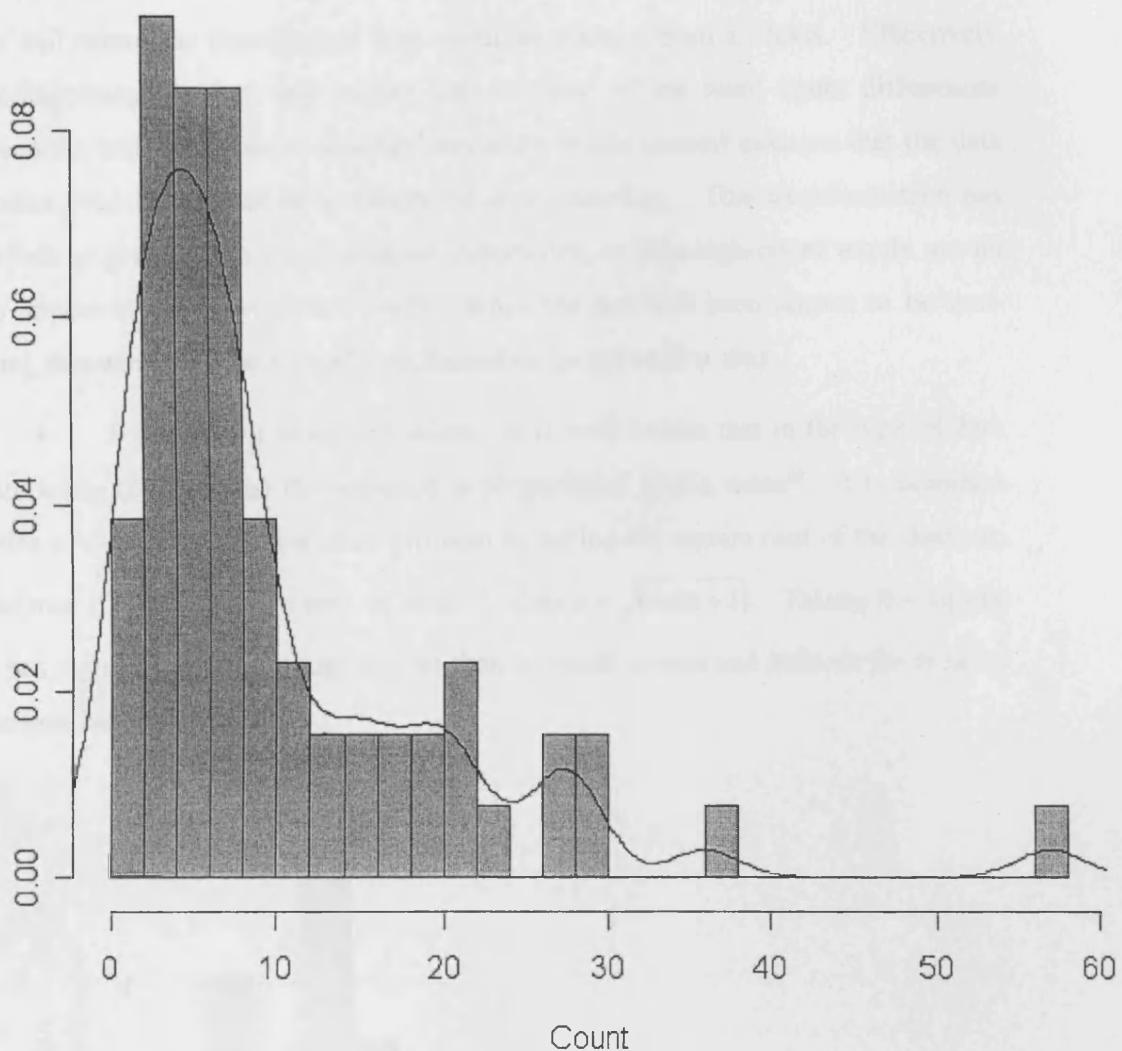
Fig. 3.1 The histogram shows that the raw data frequency of this word is highly skewed and non-continuous. The solid line is the Sheather-Jones direct plug-in density estimation of this data. The calculation of simple mean and standard deviation would be meaningless.

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,

$$Z_{rj} = \frac{x_{rj} - \mu_j}{\sigma_{jj}}$$

where μ_j = mean of column j.

σ_{jj} = standard deviation of values in column j.

This requires the mean and standard deviation of all the counts (x_r) 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[89]. 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[90], $\sqrt{data} + \sqrt{(data+1)}$. Taking the square root has a greater effect on large counts than on small counts and reduces the relative difference between them.

Transformed-Count

(a)
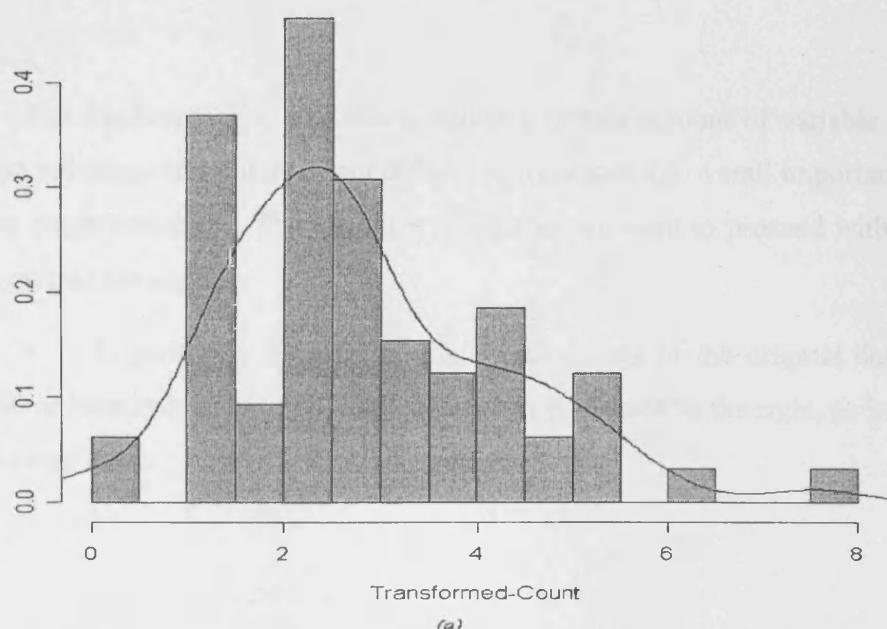
89   Bolton, S. *Pharmaceutical Statistics.* p.353

90   Weisberg, S. *Applied Linear Regression.*
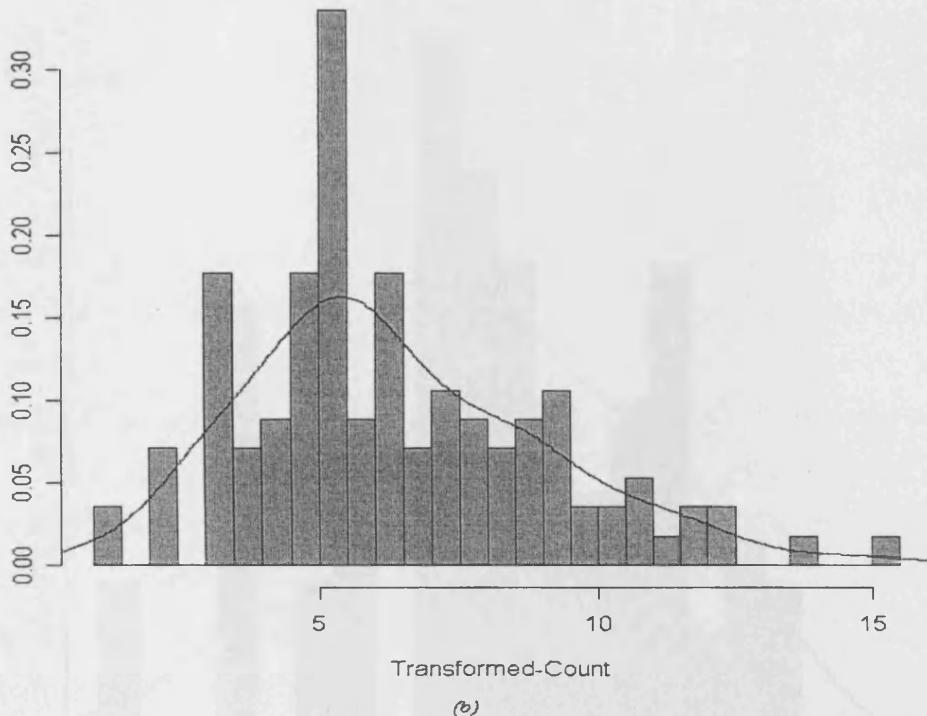
Fig. 3.2   *Frequency histogram of transformed count data of the word whose raw data*
was shown in Fig. 3.1.
(a) Simple Square Root
(b) Square Root (data) + Square Root (data + 1)

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.

Fig. 3.3 *Log transformation of the data originally shown in Fig. 3.1*

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

```
                    ┌─────────────────────┐
                    │  ALL insect groups  │
                    └──────────┬──────────┘
                               │
┌──────────────┐      ╱────────┴────────╲      ┌──────────────┐
│ ALL non-     │  No │   Does the        │ Yes │ ALL winged   │
│ winged insects│◄───│   insect have     │────►│ insects.     │
└──────────────┘      ╲   wings?        ╱      └──────────────┘
                       ╲────────────────╱
```
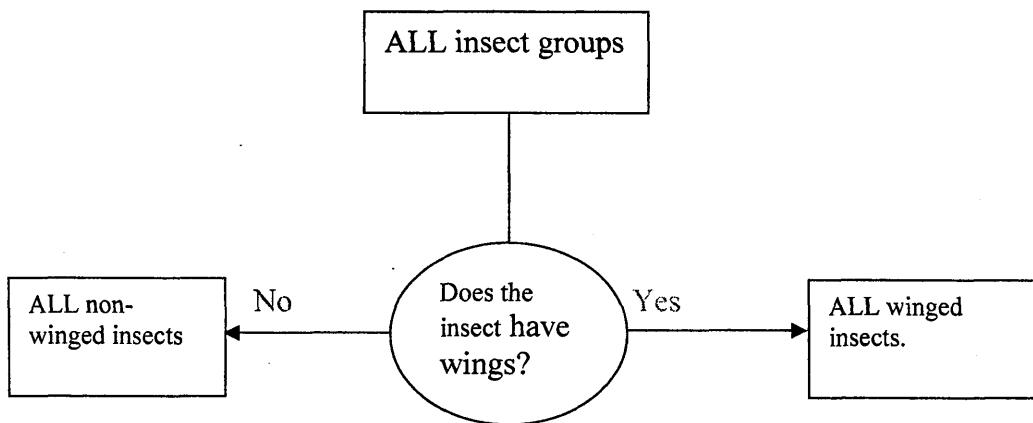
The first box diagrammatically contains all of the data. The oval represents a node, that is, a place in the tree where a partition takes place. In this instance, the node represents the constraint of insect appearance in terms of wings. There are only two possible outcomes; hence these trees are sometimes 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'.

```
                    ┌──────────────┐
                    │ ALL insect   │
                    │ groups       │
                    └──────┬───────┘
                           │
┌──────────┐        ╭──────────────╮        ┌──────────────┐
│ ALL non- │        │ Does the     │        │ ALL winged   │
│ winged   │◄───────│ insect have  │───────►│ insects.     │
│ insects  │        │ wings?       │        │              │
└────┬─────┘        ╰──────────────╯        └──────┬───────┘
     │                                             │
     ▼                                             ▼
╭──────────────╮                           ╭──────────────╮
│ Does the     │                           │ Does the     │
│ insect have a│                           │ insect have  │
│ slender body?│                           │ two pairs of │
╰──────┬───────╯                           │ wings?       │
       │╲                                  ╰──────────────╯
       ▼ ╲                                    ╱        ╲
  ╭──────────╮  ┌──────────────┐            ▼           ▼
  │ Is the   │  │  PHASMIDA    │
  │ body     │  │              │
  │grasshopper│  └──────────────┘
  │ like?    │
  ╰──┬───╲───╯
     ▼    ╲
          ┌──────────────┐
          │  ORTHOPTERA  │
          │              │
          └──────────────┘
```
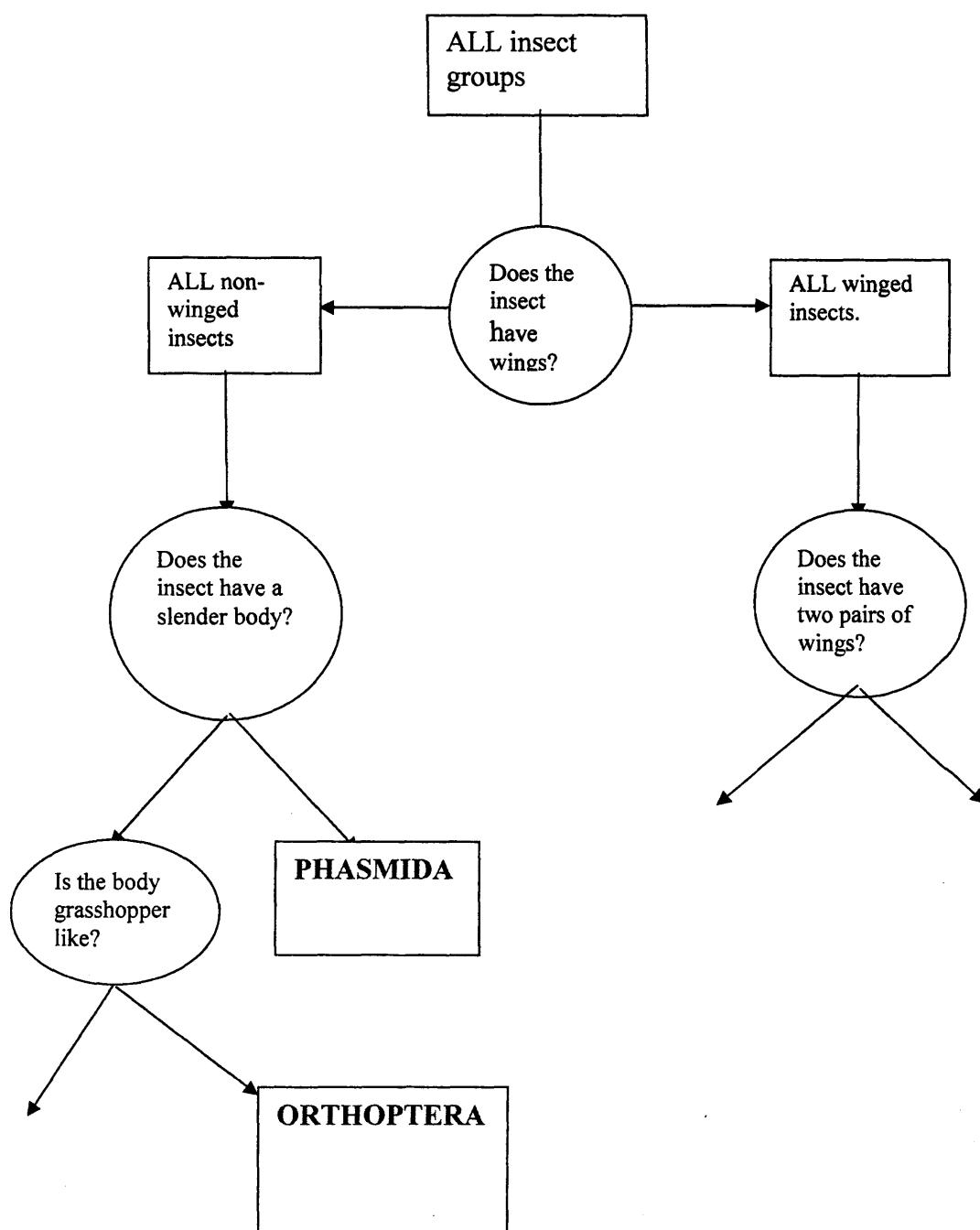
If at each node the answer to the question is yes, then follow the arrow to the right. Only part of the tree is shown here, but it can be seen that eventually the insects cannot be further partitioned. Consequently, the entire genus PHASMIDA is grouped together and branching stops. This is called a leaf 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[91] suggested a measure that would define the impurity of each node. This measure, based on "Bayesian probabilities" is called the Gini Index.

$$\text{Gini Index} = 1 - \sum_{1}^{i} \pi_i^2$$

Where π 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

---

91   Breiman, L. "Bagging Predictors". *Machine Learning* (1996) 26 No. 2, pp.123-140

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

$$I(root) = 1 - (\pi^2_{Genesis} + \pi^2_{Exodus} + ...)$$

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

$$I(root) = 0.923$$

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[92] on the raw data in supervised mode, the resulting classification tree can be grown:

---

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.cfmin.

Count 99>=0.5

Count 134< 14.5   Count 52< 2

Count 95< 4.5   Count 179>=38.5   Count 16< 0.5

Count 25>=1 .Count < 0.5 G   Ex   Si

2Es G   Ex IIC

Count 214>=5.5

Jn

Count 180>=0.5

A

Count 129< 3.5

Lk

Count 53< 35   Count 0>=3

Count 7>=1.5   M   S

Count < 1.5   Rv

IC   Count 27< 0.5

Count 79>=17   Ro

Count 188>=0.5   So

Count 188>=0.5
Count 4< 4   V
2Es
G   Ex
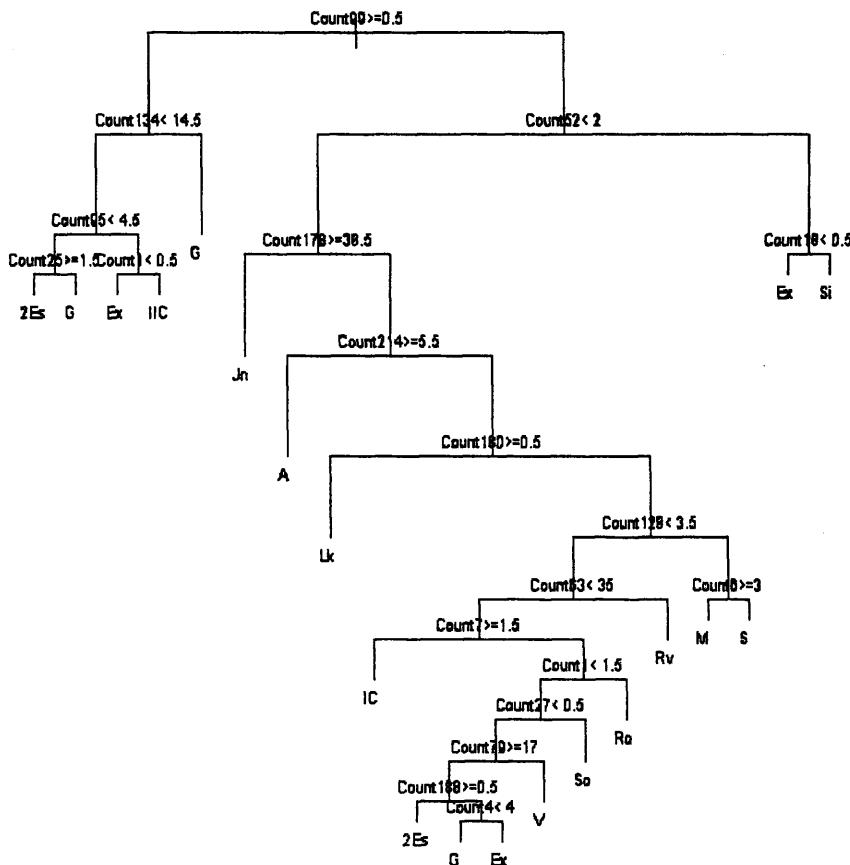
Fig.4.1 An example of a single binary partition tree based on the original data contained in a 62x230 data matrix. The tree is un-pruned.

KEY					
2Es	2 Esdras	IIC	2 Corinthians	Si	Sirach
A	Acts	Jn	John	Si	Similitudes
Ex	Exodus	Lk	Luke	So	Wisdom
G	Genesis	M	Mandates		
IC	1 Corinthians	Ro	Romans		
V	Visions	Rv	Revelation		

The root node is first split using the constraint Count 99>= 0.5   to the right, that is, if the count of the word ἡνίκα (a particle that expresses time, 'when', 'the time when') is greater than, or equal to, 0.5 then take the right branch etc.   ἡνίκα 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 ie. 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:

Unsupervised Random Forest Variable Importance
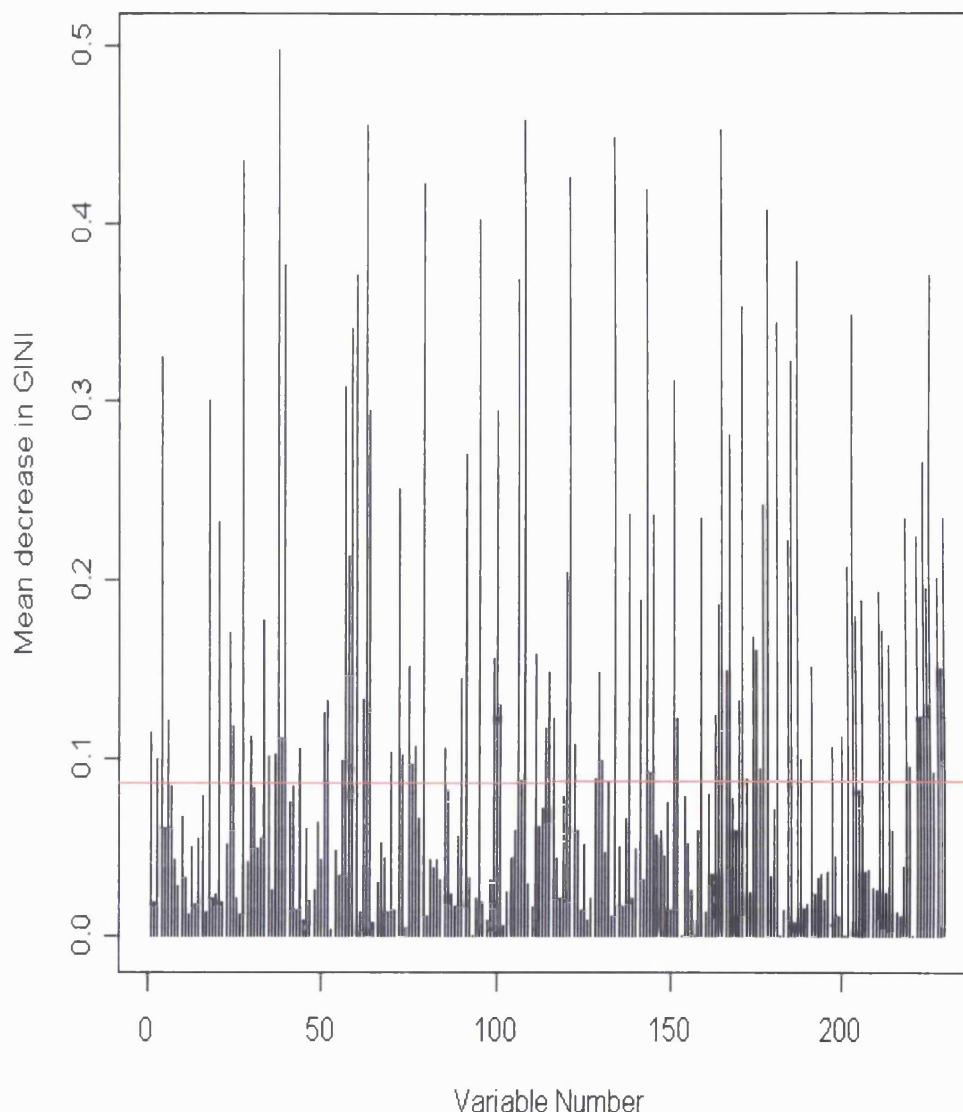
<u>**Fig.4.2**</u>   *The figure shows the relative importance of each variable count used in the analysis. The greater the mean decrease in GINI, the more significant the variable count. The red line shows a suggested cut-off point. Variables below the red line are not of significant importance in discrimination. 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

69

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:

randomForest(formula = data.df1[, 1] ~ ., data = data.df1[, -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:

	2Es	A	Ex	G	IC	IIC	Jn	Lk	M	Ro	Rv	S	Si	So	V	class.error
2Es	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.000
A	0	4	0	0	0	0	0	1	0	0	0	0	0	0	0	0.200
Ex	0	0	5	3	0	0	0	0	0	0	0	0	0	0	0	0.375
G	0	0	1	9	0	0	0	0	0	0	0	0	0	0	0	0.100
IC	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0.500
IIC	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1.000
Jn	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0.000
Lk	0	1	0	0	0	0	0	5	0	0	0	0	0	0	0	0.167
M	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	1.000
Ro	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0.500
Rv	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0.000
S	0	0	0	0	0	0	0	0	1	0	0	3	0	0	0	0.250
Si	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0.000
So	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0.500
V	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	1.000

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 Lukan Corpus shows Luke has one of its cases in the Acts column the remainder being classified as Luke. Similarly with Acts. Therefore, Luke-Acts have parts in common.

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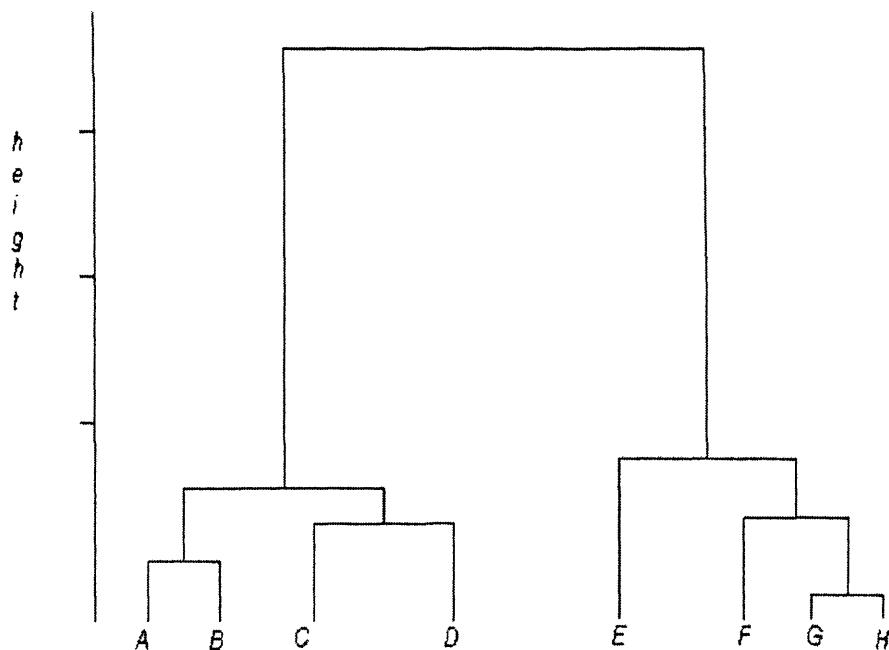
Fig.5.1 A stylised dendrogram showing the structure of nodes and branches.

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:
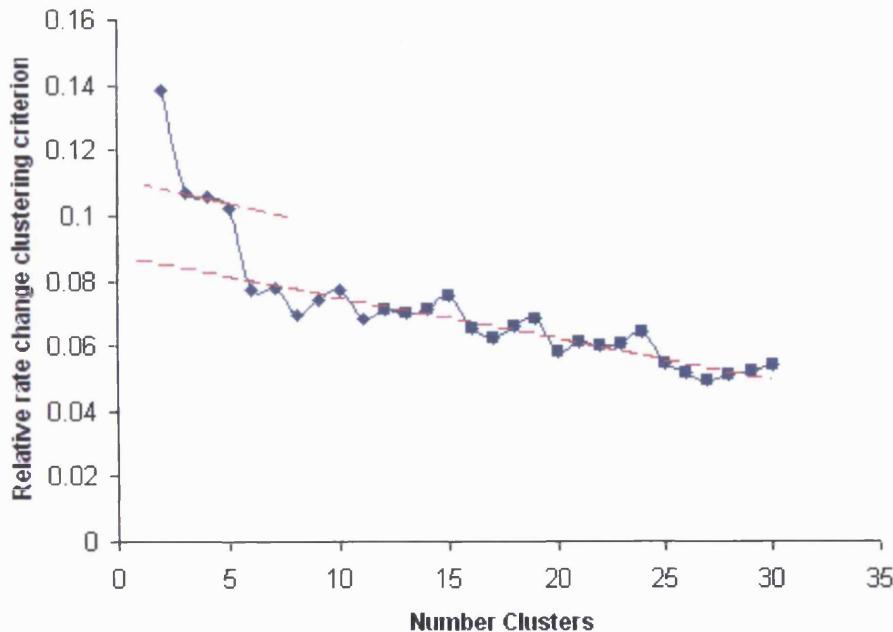
Fig. 5.2   *Typical rate of change of clustering criterion plot*

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:

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

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

$$W_1 = \sum_{r=1}^{C1} \sum_{q=1}^{nr} \left(x_{rq} - \bar{x}_r\right)' \left(x_{rq} - \bar{x}_r\right)$$

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 $\sqrt{count} + \sqrt{(count + 1)}$ 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; $F^* = 2.65$, $F_{crit} = 2.54$ 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.

Dendrogram of Z-transformed Euclidean Distance
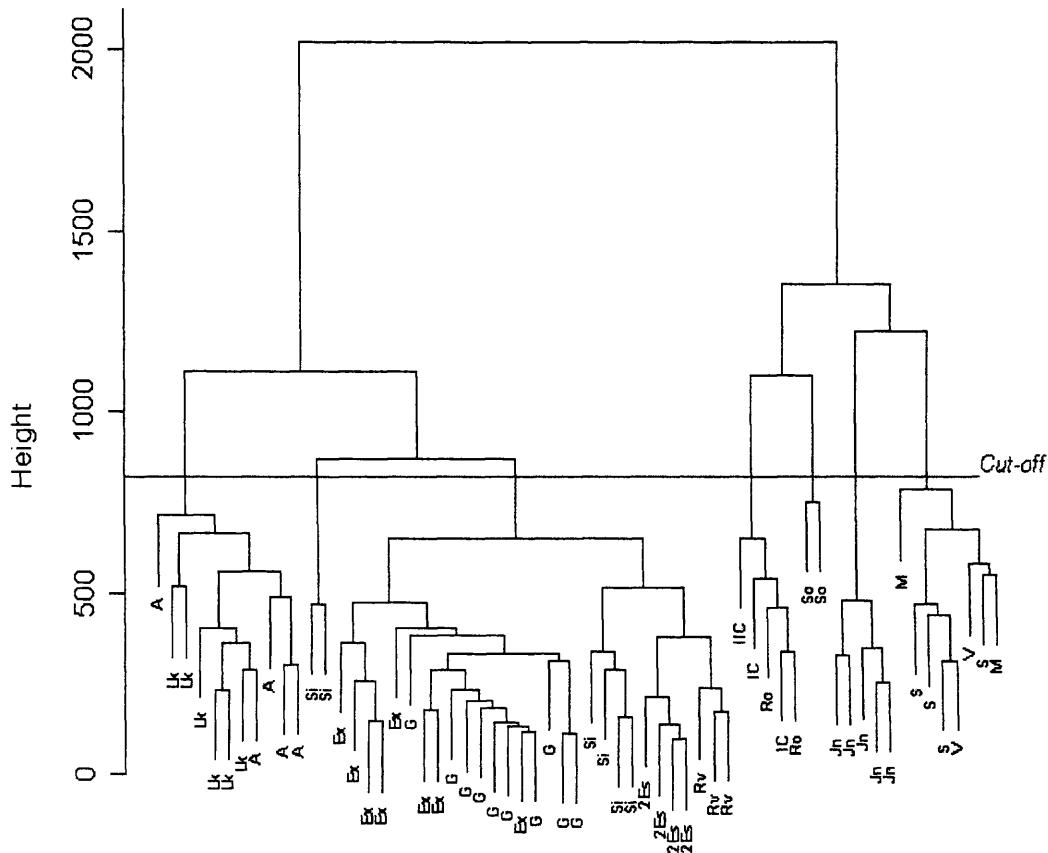
Fig.5.3   The dendrogram was formed from the standardized values of the 62x230 data matrix of word counts. The squared Euclidean distance was clustered using Ward's method.

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 andActs, for example, are clustered together.

KEY					
2Es	2 Esdras	Jn	John	Si	Similitudes
A	Acts	Lk	Luke	So	Wisdom
Ex	Exodus	M	Mandates	V	Visions
G	Genesis	Ro	Romans		
IC	1 Corinthians	Rv	Revelation		
IIC	2 Corinthians	Si	Sirach		

Z-transformed Data
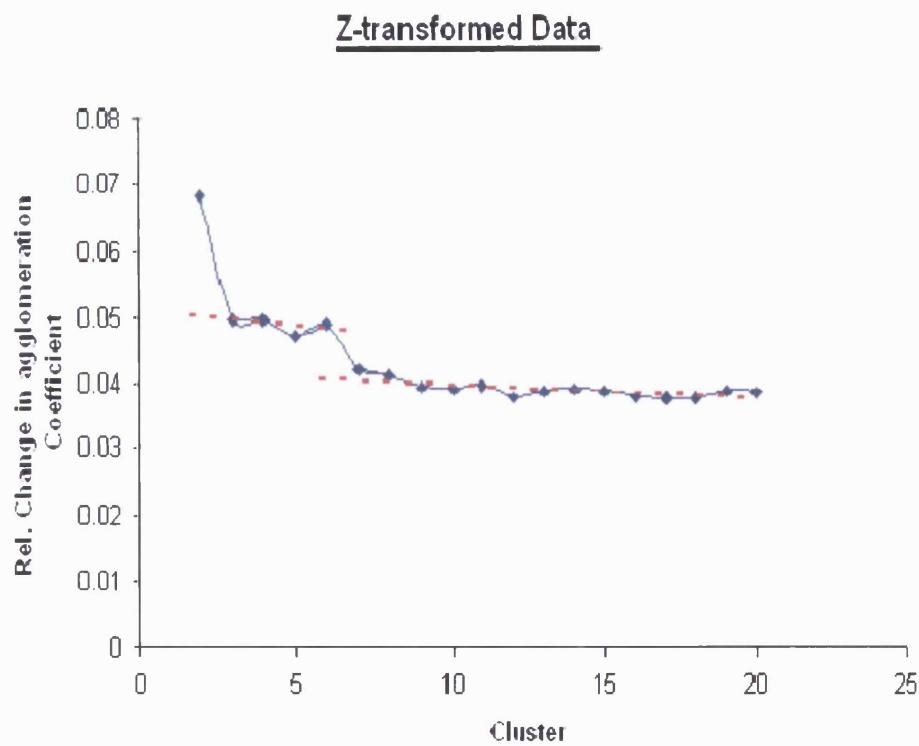
Fig.5.4 *Relative change in the agglomeration can be seen to rise when going from 7 clusters to 6. The other change occurs in going from 3 clusters to 2. Trend lines are shown in red.*

The cases are partitioned into the 7 clusters as follows:

	1	2	3	4	5	6	7
2Es	4	0	0	0	0	0	0
A	0	0	0	5	0	0	0
Ex	8	0	0	0	0	0	0
G	10	0	0	0	0	0	0
IC	0	0	0	0	2	0	0
IIC	0	0	0	0	1	0	0
Jn	0	5	0	0	0	0	0
Lk	0	0	0	6	0	0	0
M	0	0	0	0	0	0	2
Ro	0	0	0	0	2	0	0
Rv	3	0	0	0	0	0	0
S	0	0	0	0	0	0	4
Si	4	0	0	0	0	2	0
So	0	0	2	0	0	0	0
V	0	0	0	0	0	0	2

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.

Dendrogram of SQRT-transformed Euclidean Distance
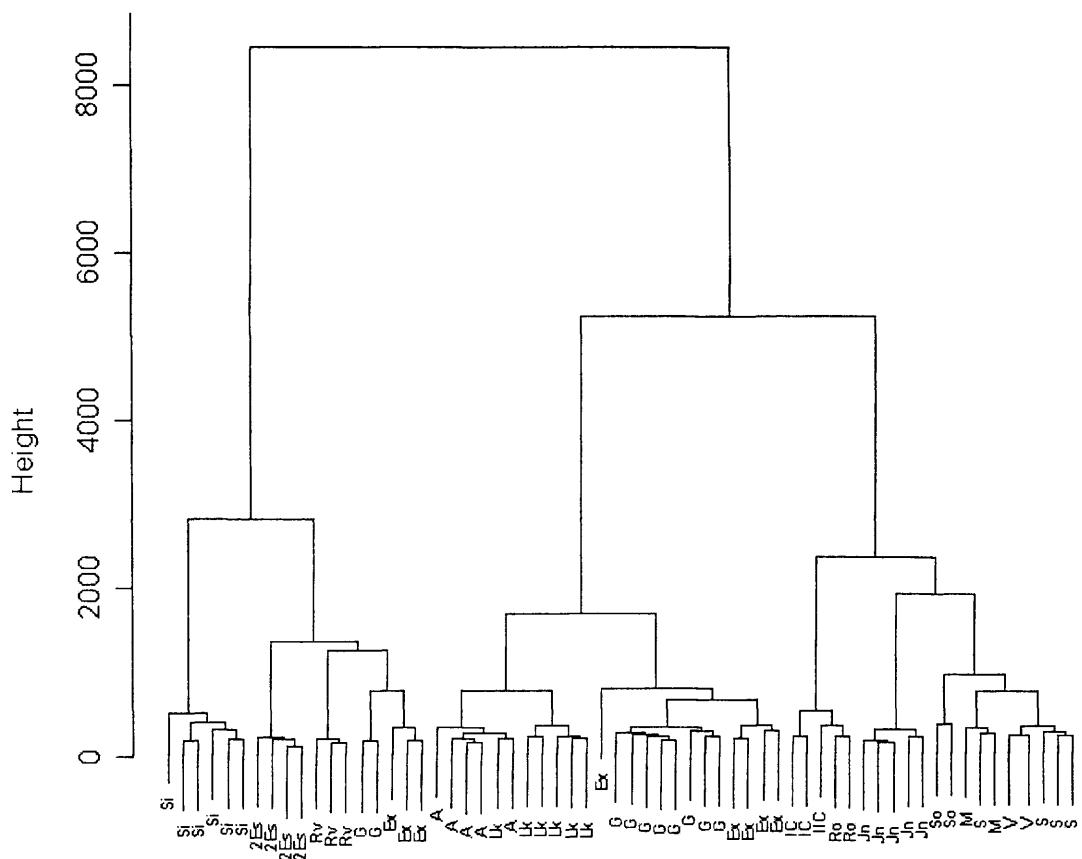
Fig.5.5 Dendrogram obtained from 60x230 matrix after values had been transformed by SQRT(data)+SQRT(data+1), followed by Ward's clustering method on the squared euclidean distance transformation.

81

Change of Agglomeration Coefficient SQRT-transformed Data
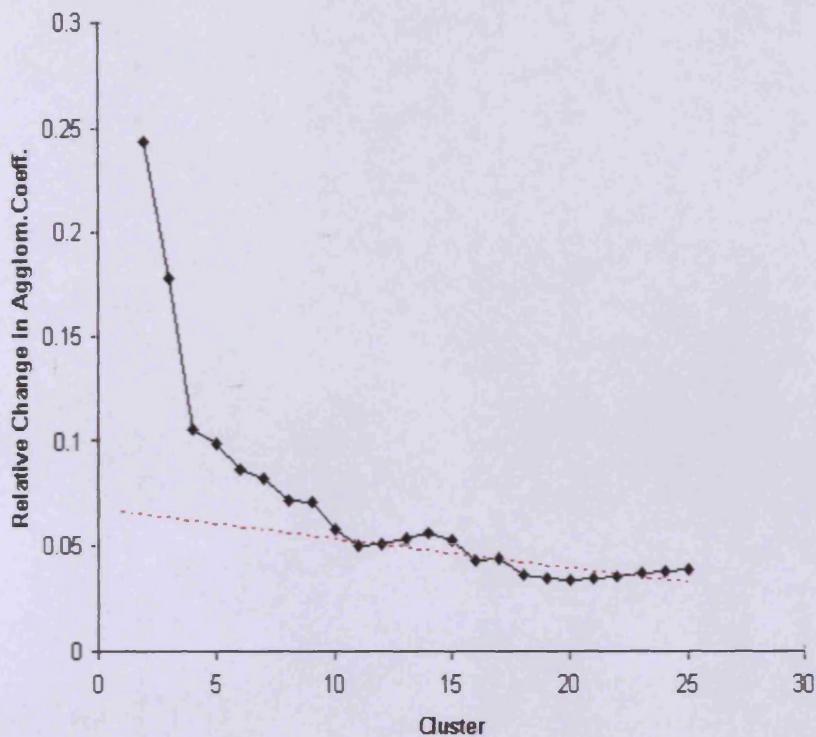
Fig.5.6 *Relative change in the agglomeration coefficient shows a slight increase in going from 16 to 15 clusters and 11 to 9 clusters. Trend line is shown in red.*

The cases are partitioned into the 11 clusters as shown in the table below.

	1	2	3	4	5	6	7	8	9	10	11
2Es	4	0	0	0	0	0	0	0	0	0	0
A	0	0	0	0	0	0	5	0	0	0	0
Ex	0	0	3	1	4	0	0	0	0	0	0
G	0	0	2	0	8	0	0	0	0	0	0
IC	0	0	0	0	0	0	0	2	0	0	0
IIC	0	0	0	0	0	0	0	1	0	0	0
Jn	0	5	0	0	0	0	0	0	0	0	0
Lk	0	0	0	0	0	0	6	0	0	0	0
M	0	0	0	0	0	0	0	0	0	2	0
Ro	0	0	0	0	0	0	0	2	0	0	0
Rv	0	0	0	0	0	0	0	0	0	0	3
S	0	0	0	0	0	0	0	0	0	4	0
Si	0	0	0	0	0	0	0	0	6	0	0
So	0	0	0	0	0	2	0	0	0	0	0
V	0	0	0	0	0	0	0	0	0	2	0

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.

Dendrogram from RF-Proximity Matrix
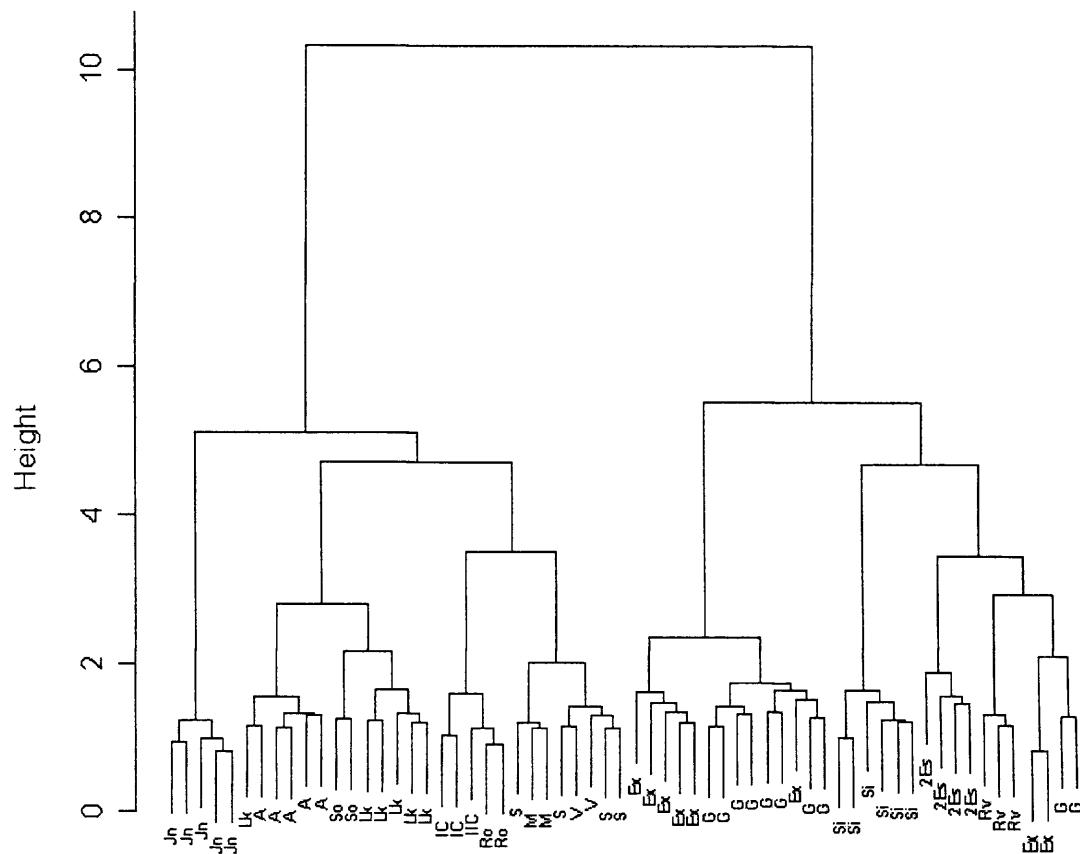
Fig.5.7 _Dendrogram produced by the RF-proximity matrix clustered using Ward's method._

Change of Agglomeration Coefficient RF-Proximity Data
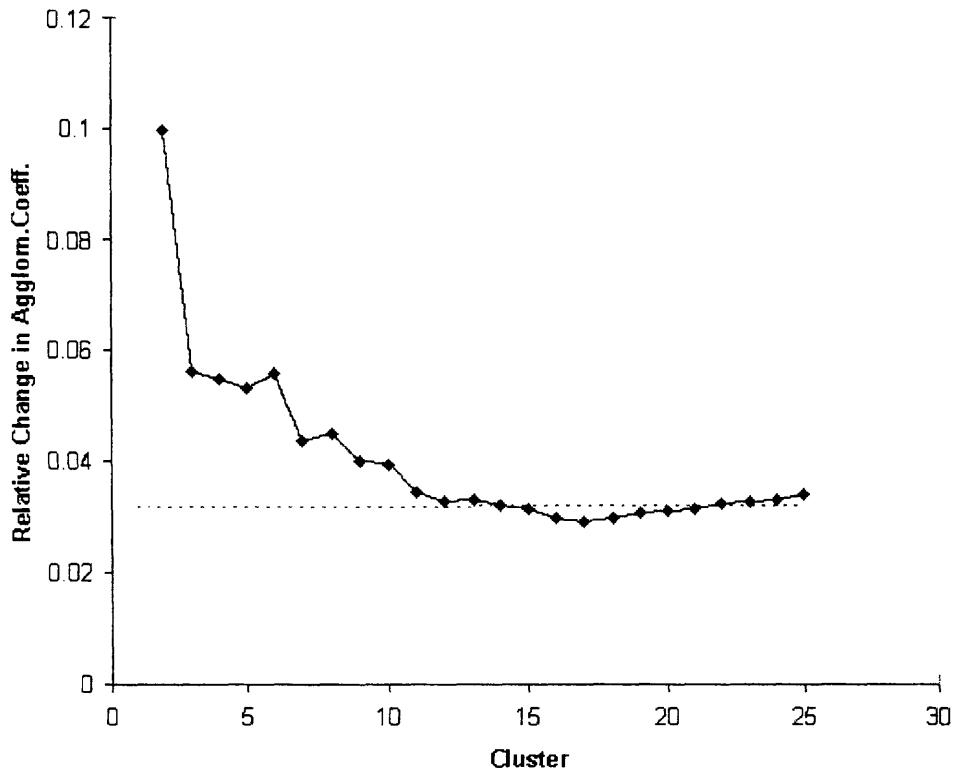
Fig.5.8   Relative change in the agglomeration coefficient suggests a clustering solution
of 11, 10 or 7 clusters.

Beale's Pseudo F at 95% confidence favours a 7-cluster solution. The cases are partitioned as shown in the table below:

	1	2	3	4	5	6	7
2Es	4	0	0	0	0	0	0
A	0	0	0	5	0	0	0
Ex	2	0	6	0	0	0	0
G	2	0	8	0	0	0	0
IC	0	0	0	0	2	0	0
IIC	0	0	0	0	1	0	0
Jn	0	5	0	0	0	0	0
Lk	0	0	0	6	0	0	0
M	0	0	0	0	0	0	2
Ro	0	0	0	0	2	0	0
Rv	3	0	0	0	0	0	0
S	0	0	0	0	0	0	4
Si	0	0	0	0	0	6	0
So	0	0	0	2	0	0	0
V	0	0	0	0	0	0	2

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.
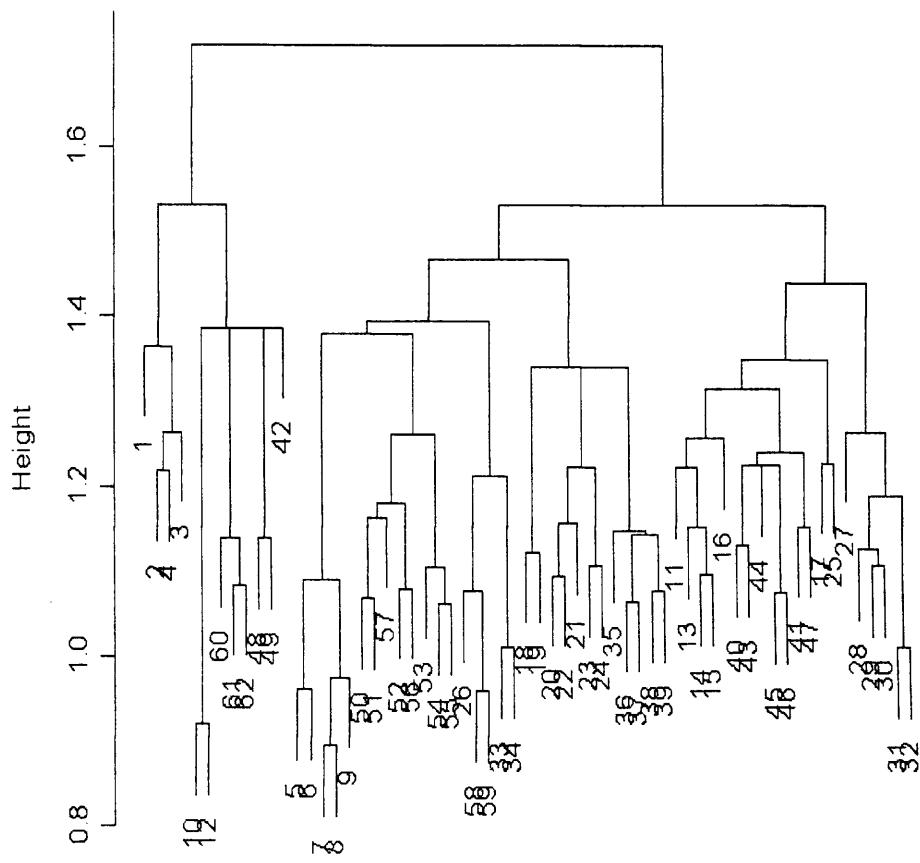
Fig.5.9 Dendrogram of RF proximity matrix by divisive clustering is more difficult to interpret because of its 'bushy' appearance. Leaf node labels refer to the texts used in the analysis.

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

RF Proximity Divisive Method
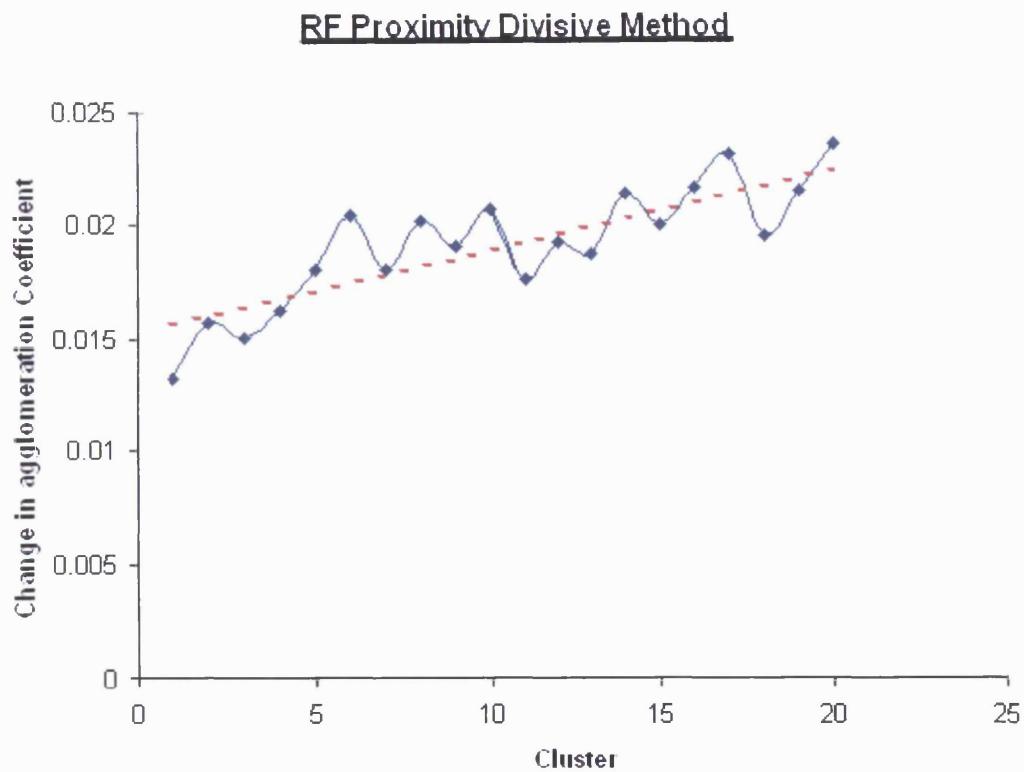
Fig.5.10   The trend line (in red) is based on the last 10 data points, but because the
fluctuations in agglomeration coefficient are large, the trend line could cover
the whole of the data, making discovery of an optimal clustering impossible.

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.

88

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 index[96] 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) = 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
96   Kaufman,L., Rousseeuw,P.J. *Finding Groups in Data. An Introduction To Cluster Analysis.*

Average Silhouette Width

Cluster

(a)
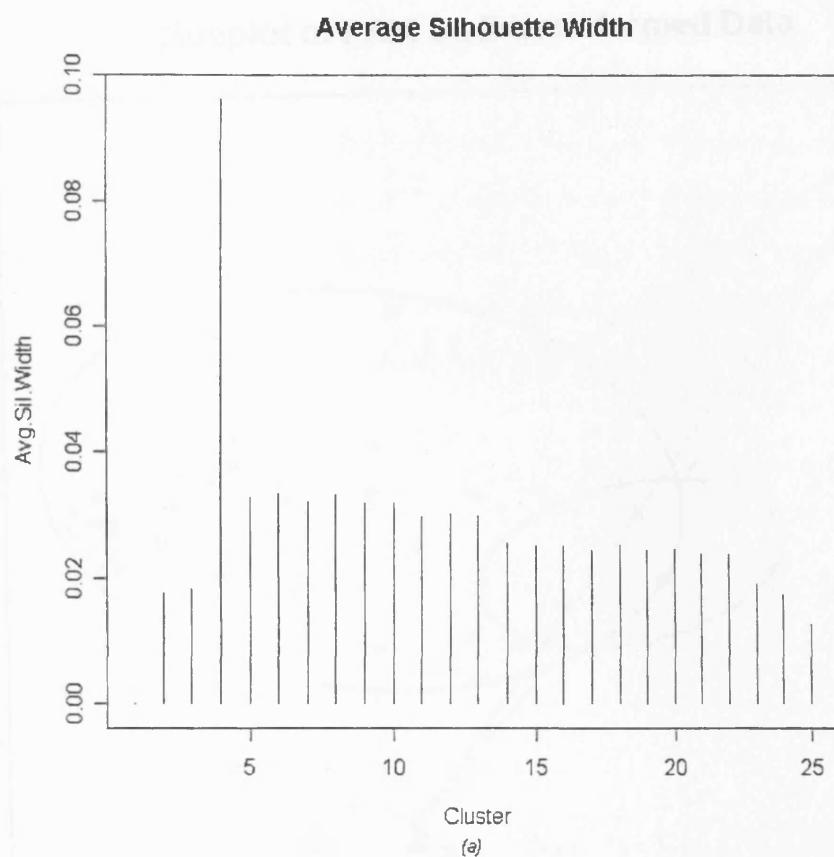
Silhouette Plot of Z-Transformed Data Using PAM with 4 Clusters

n = 62

4 clusters C_j
$j: n_j \mid ave_{i \in C_j} \ s_i$

1: 51 | 0.10

2: 5 | 0.15
3: 1 | 0.00
4: 5 | 0.07

Silhouette width s_i

Average silhouette width: 0.1   *(b)*

90
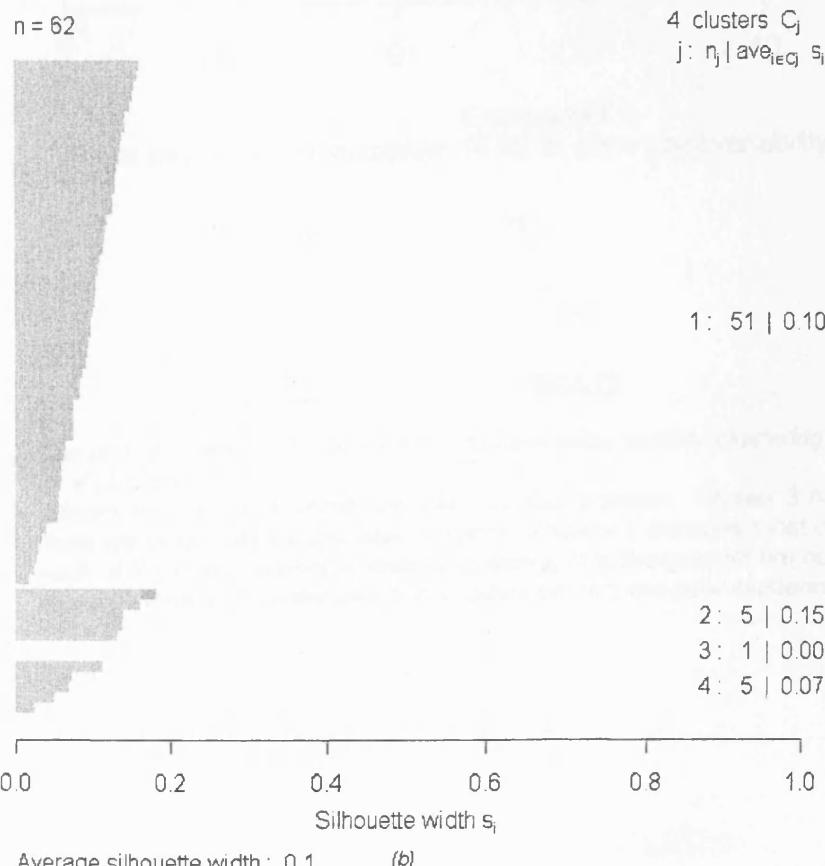
Clusplot of PAM on Z-transformed Data

Component 1

These two components explain 16.92 % of the point variability.

(c)
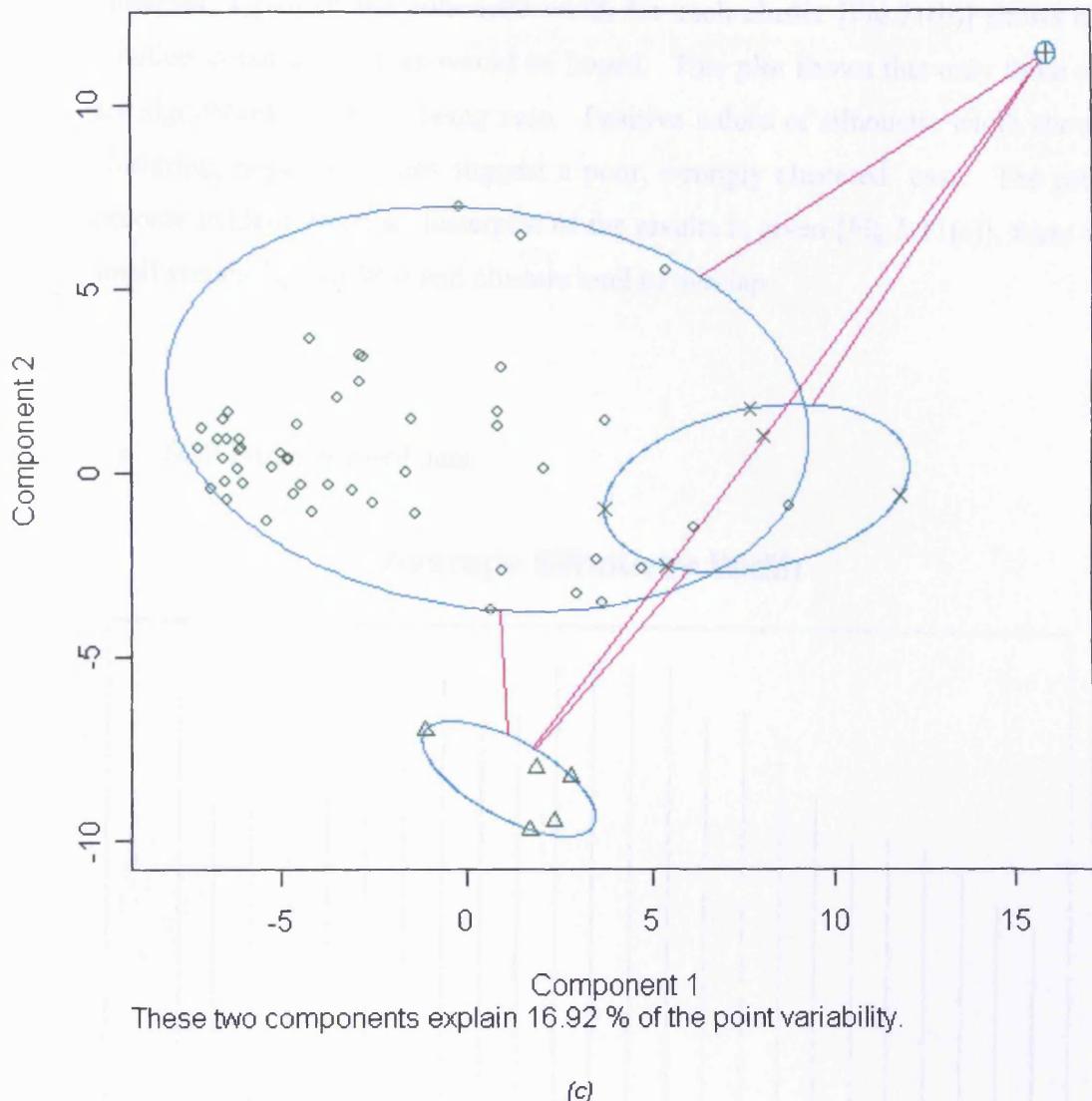
Fig.5.11

(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.Width 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.

91

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.51(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.

Average Silhouette Width

(a)
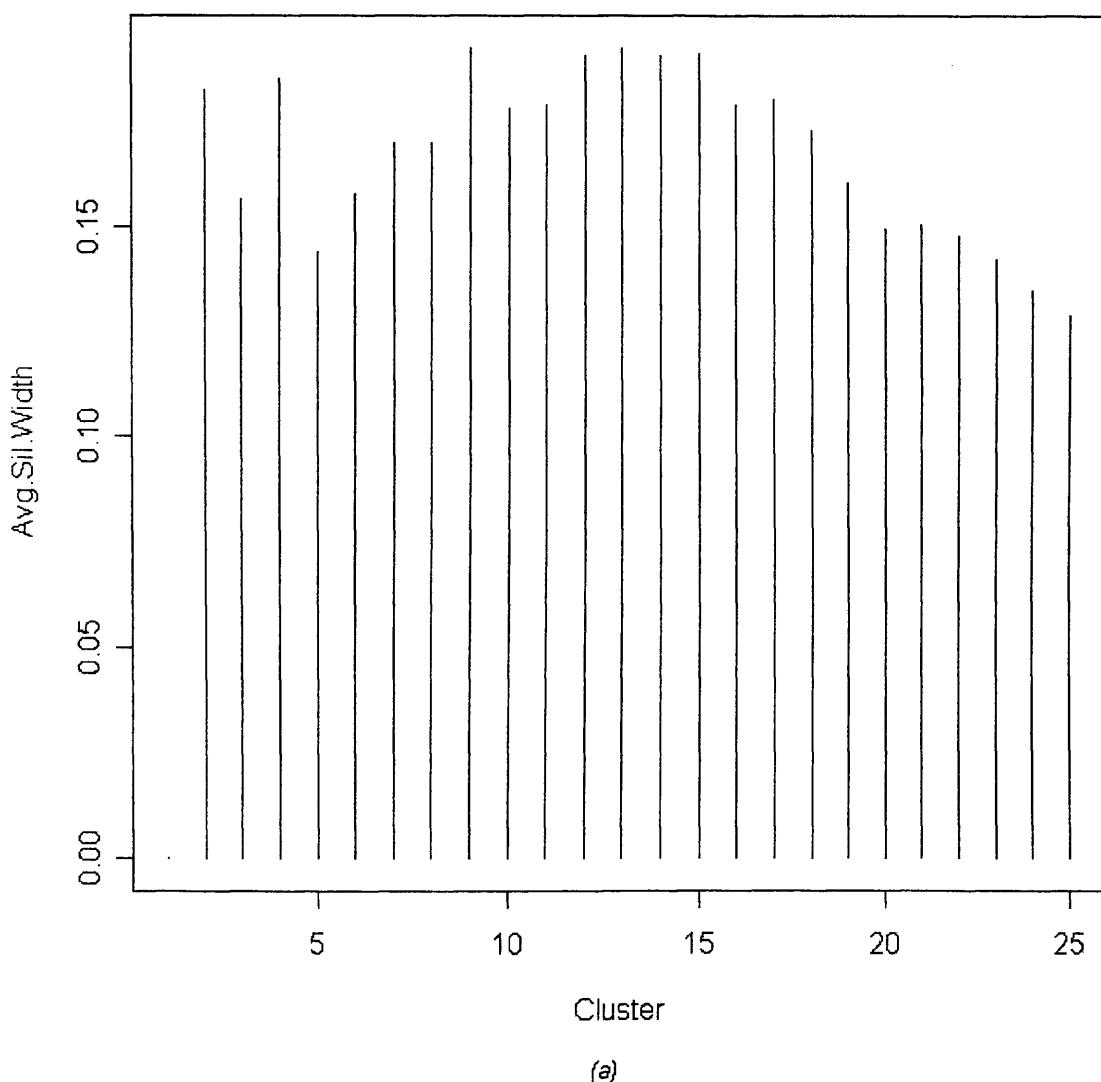
Silhouette Plot Of SQRT-transformed Data Showing 9 Clusters

n = 62
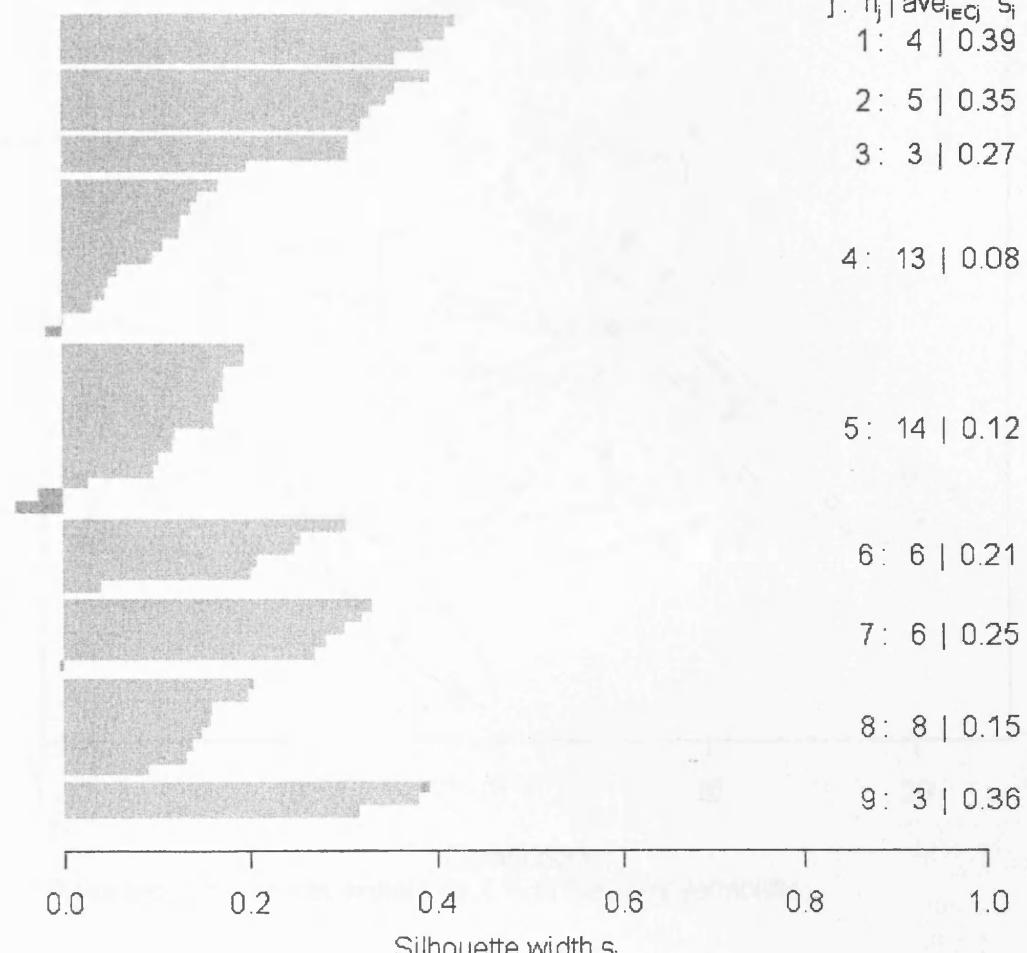
9 clusters C_j

$j:\ n_j\ |\ ave_{i \in C_j}\ s_i$

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

Silhouette width s_i

Average silhouette width : 0.19

(b)

Clusplot SQRT-transformed Data

These two components explain 41.1 % of the point variability.

(c)
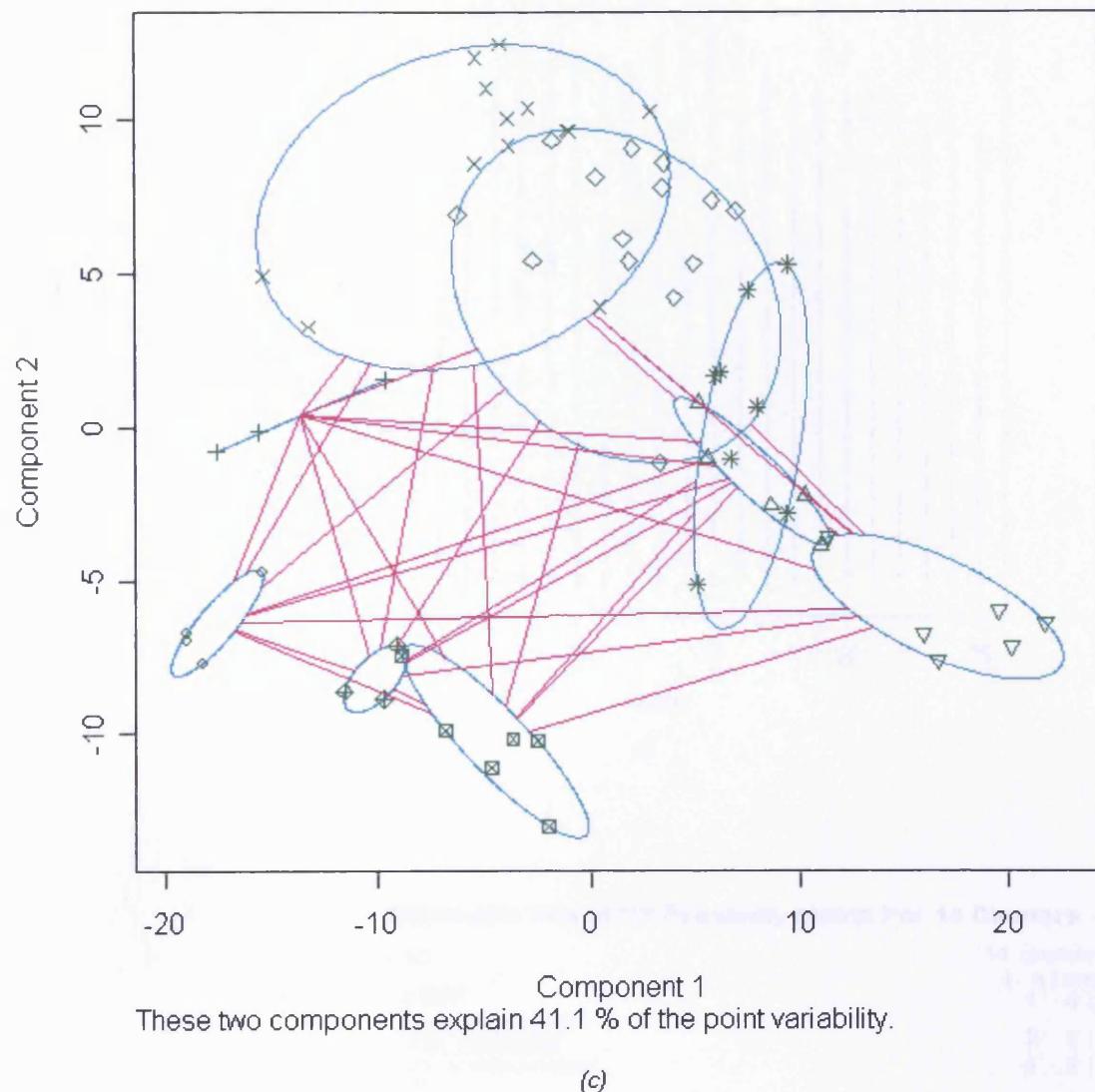
Fig.5.12

(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

(a)
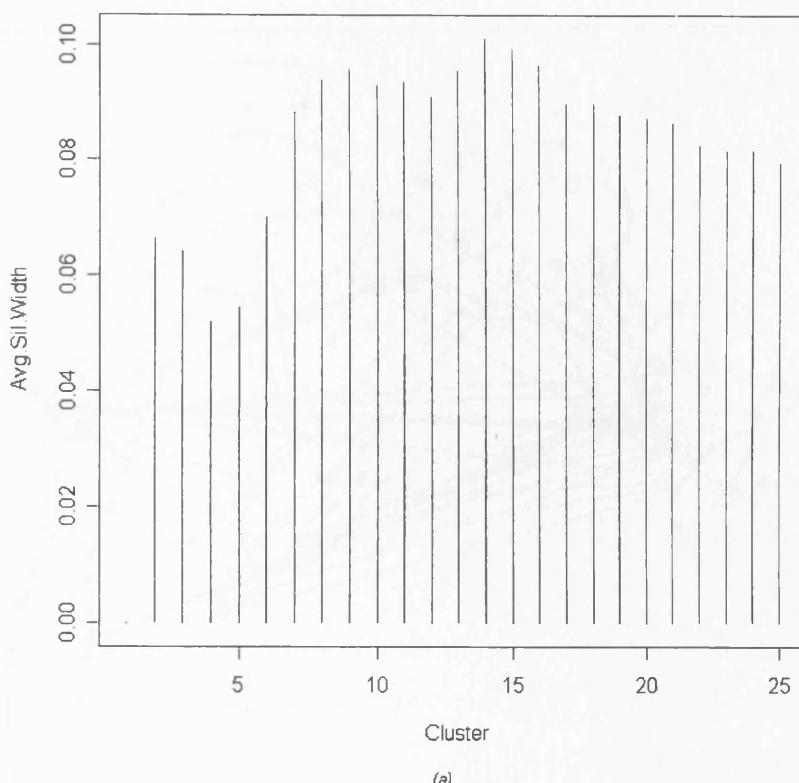
Silhouette Plot of RF-Proximity Matrix For 14 Clusters

n = 62

| 14 clusters C_j |
j: n_j | $ave_{i \in C_j}$ s_i
1: 4 | 0.07
2: 5 | 0.23
3: 2 | 0.27
4: 7 | 0.02
5: 2 | 0.11
6: 4 | 0.08
7: 7 | 0.07
8: 5 | 0.14
9: 6 | 0.13
10: 7 | 0.06
11: 2 | 0.10
12: 5 | 0.06
13: 3 | 0.10
14: 3 | 0.16

Silhouette width s_i

Average silhouette width : 0.1

(b)
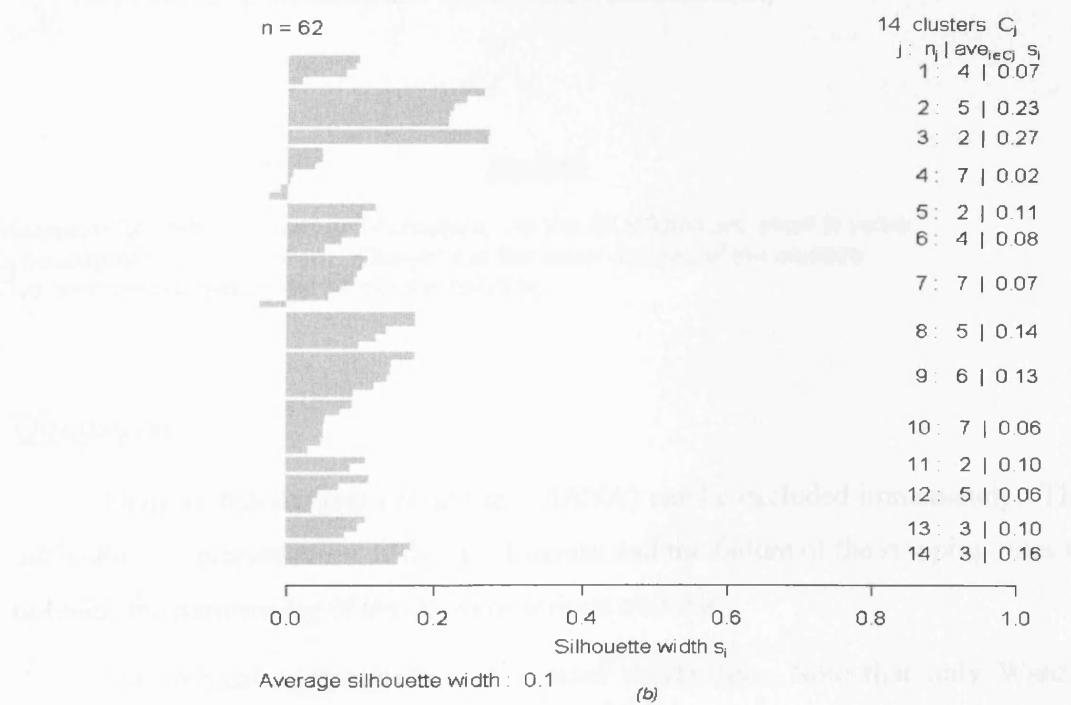
Clusplot For RF-Proximity Matrix (14 clusters)
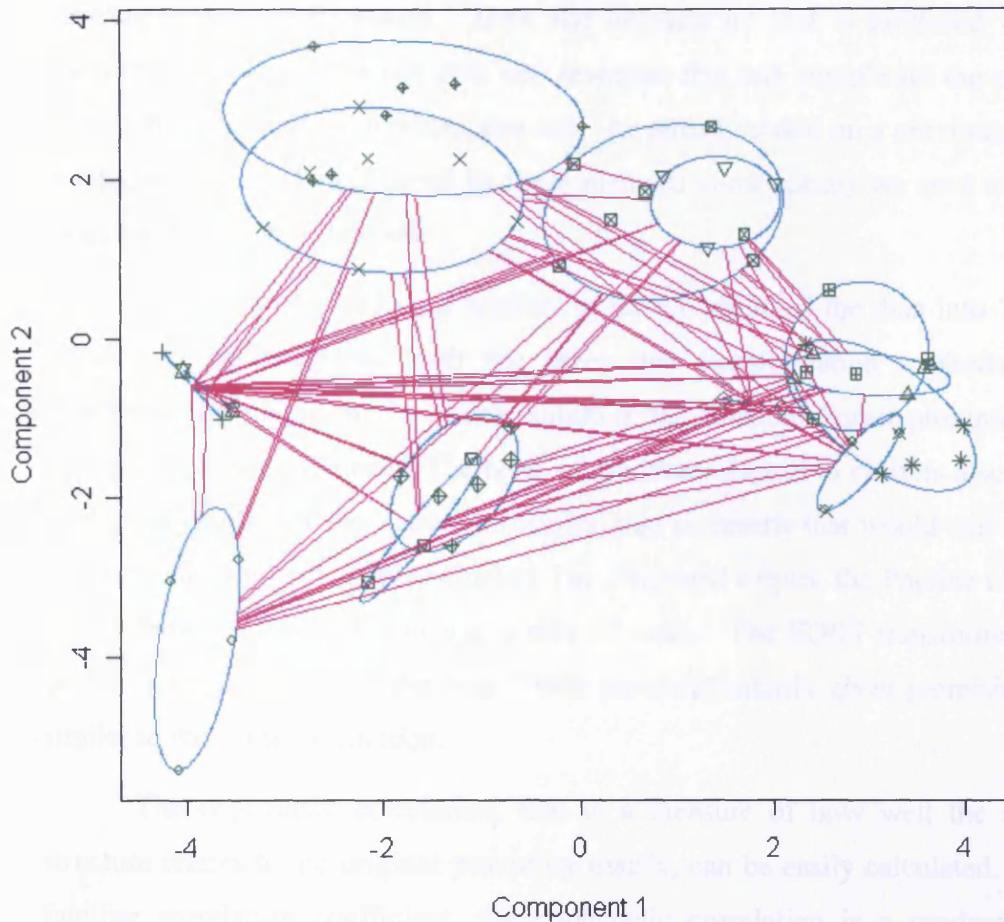
Component 1
These two components explain 16.72 % of the point variability.

(c)

Fig.5.13

(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[97].

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

---

97   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 ie. 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:

$$\frac{\left(c_i' d \right)^2}{\lambda_i}$$

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:

καί	δέ	οὐ, οὐχ, οὐκ	ἐπί, ἐπ᾽, ἐφ᾽	μή
ὅτι	ἐκ, ἐξ	γάρ	ἀπό, ἀπ᾽, ἀφ᾽	πρός
μετά, μετ᾽, μεθ᾽	διά, δι᾽	ἵνα, ἵν᾽	ὡς	ἀλλά, ἀλλ᾽
ἐάν, ἐάνπερ	κατά, κατ᾽, καθ᾽	εἰ, εἴπερ, εἴπως	περί	ἤ, ἤτοι
παρά, παρ᾽	οὕτω, οὕτως	ὑπό, ὑπ᾽, ὑφ᾽	νῦν, νυνί	πάντως
μέν	ἐκεῖ, ἐκεῖθεν	ἄν	ἕως	ἔτι
οὐδέ	καθώς	ὑπέρ	ὅταν	πάλιν
ἐνώπιον	οὔτε	πῶς	ἀνά	σύν
ὥστε				

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.

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
Eigenvalue	8.95	3.45	2.88	2.39	2.18
% Variance	21.84	8.42	7.03	5.83	5.31
Cumulative %	21.84	30.25	37.28	43.11	48.42

	Comp.6	Comp.7	Comp.8	Comp.9	Comp.10
Eigenvalue	2.00	1.64	1.63	1.36	1.28
% Variance	4.89	4.01	3.97	3.32	3.13
Cumulative %	53.31	57.32	61.29	64.62	67.75

	Comp.11	Comp.12	Comp.13	Comp.14	Comp.15
Eigenvalue	1.08	0.97	0.93	0.85	0.80
% Variance	2.63	2.36	2.27	2.08	1.94
Cumulative %	70.38	72.74	75.01	77.09	79.03

The associated scree plot is shown for completion.

Scree Plot Z-Transformed Data
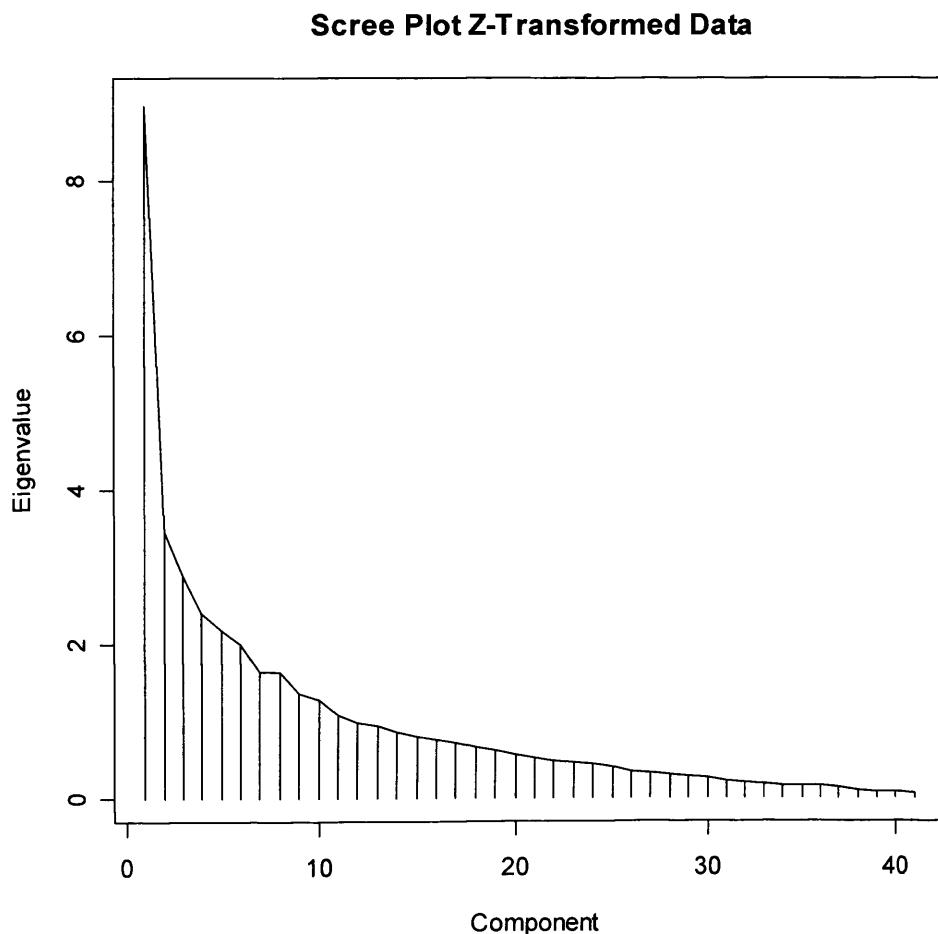
Fig.6.1 Scree Plot of eigenvalues against the number of components from PCA. Components with eigenvalues less than 1 are not significant. In this case the first 11 components are significant.

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 is 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
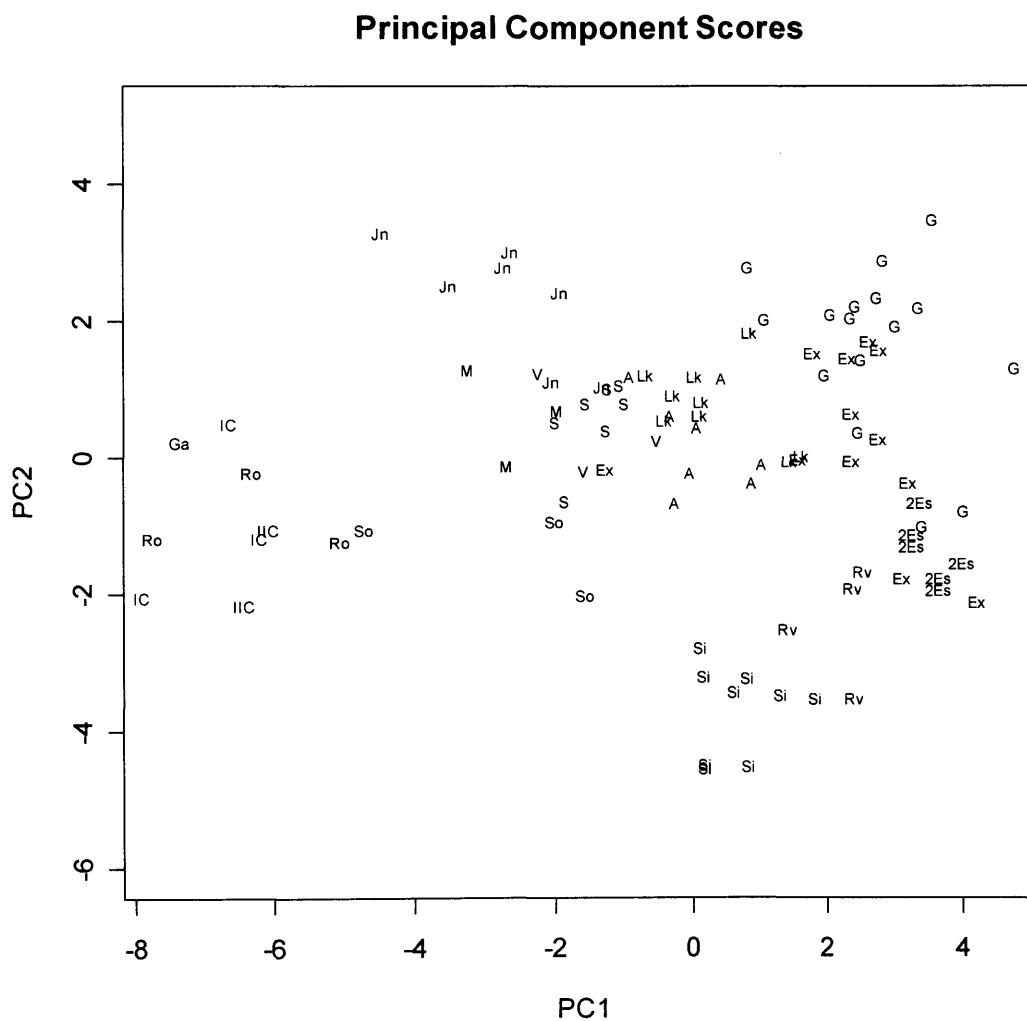
Fig.6.2

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.
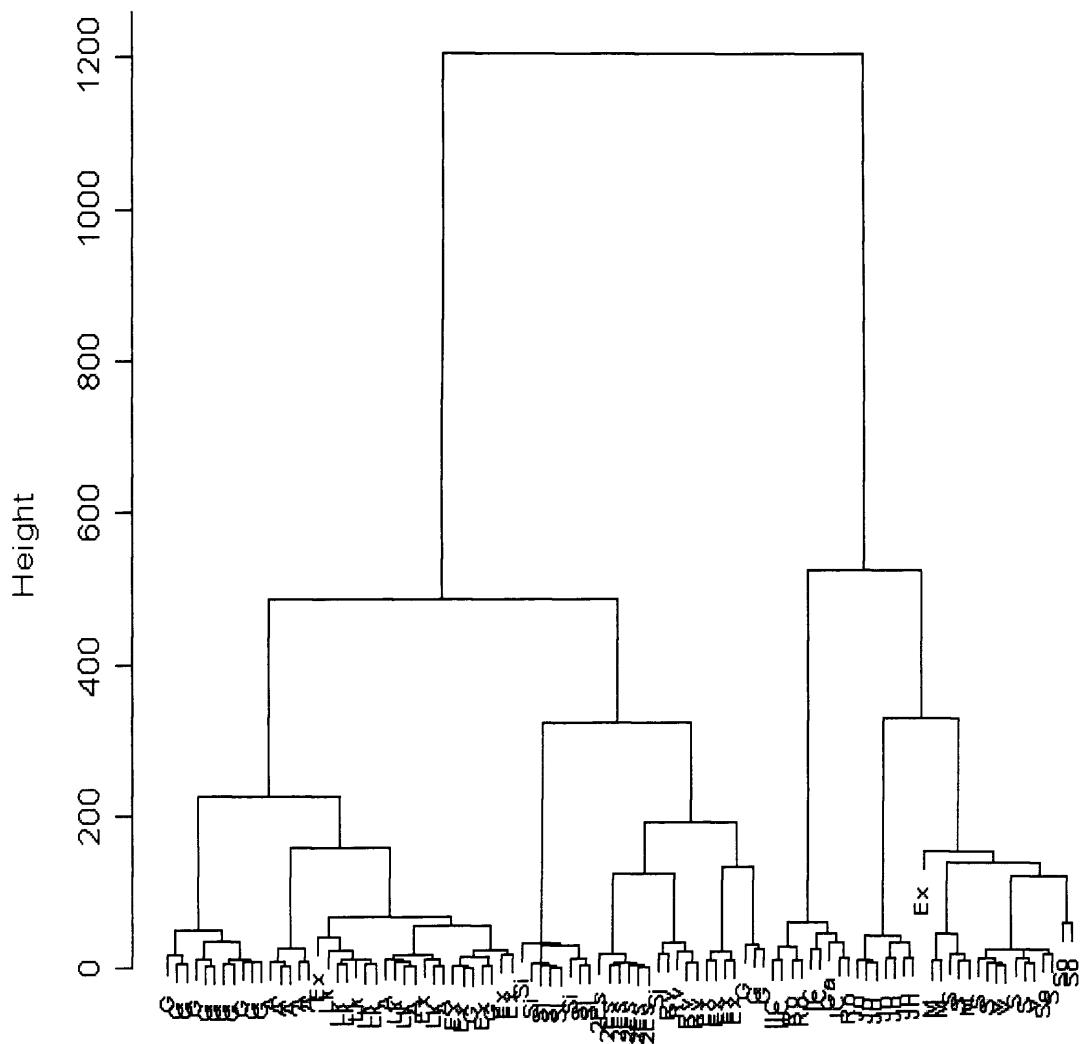
Fig.6.3 Dendrogram of the first 11 Principal Components by Ward's Method. It shows a clear clustering of the Pauline corpus.

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 Change in Agglomeration Coefficient
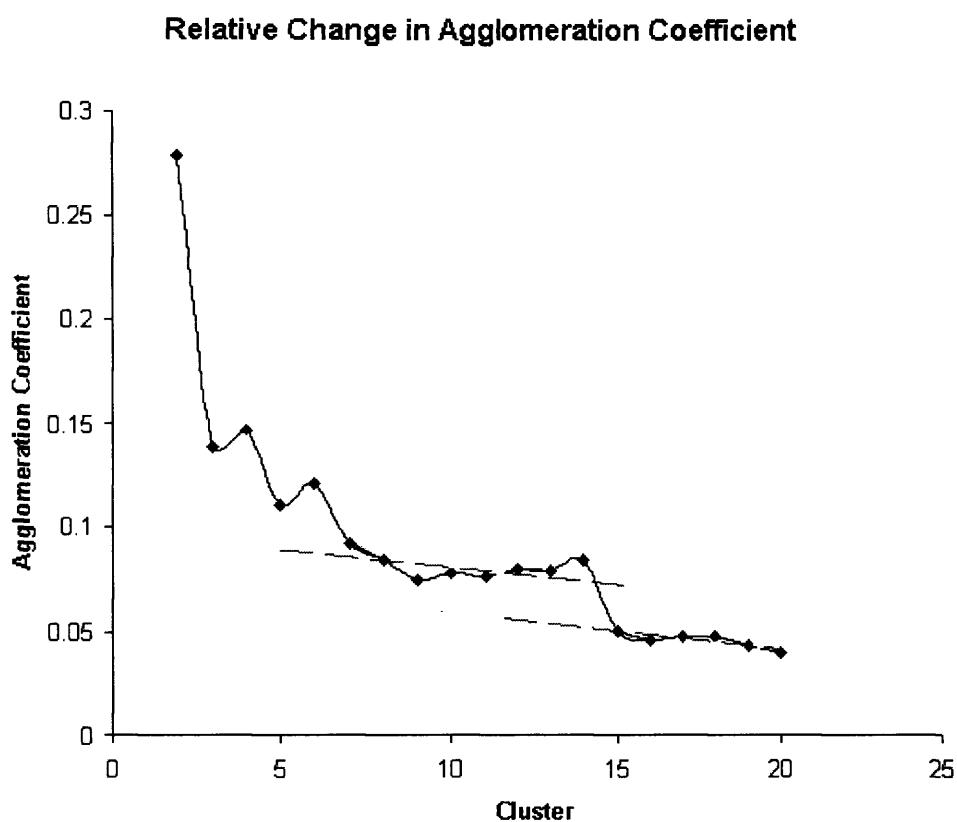
Fig.6.4 Relative changes in the agglomeration coefficient can be noted in going from 15 to 14 clusters, 8 to 7 clusters and from 5 to 4 clusters.

Cluster	1	2	3	4	5	6	7
2Es	6	0	0	0	0	0	0
A	0	0	8	0	0	0	0
Ex	4	0	7	1	0	0	0
G	3	0	2	0	0	0	11
Ga	0	0	0	0	1	0	0
IC	0	0	0	0	3	0	0
IIC	0	0	0	0	2	0	0
Jn	0	7	0	0	0	0	0
Lk	0	0	9	0	0	0	0
M	0	0	0	3	0	0	0
Ro	0	0	0	0	3	0	0
Rv	4	0	0	0	0	0	0
S	0	0	0	7	0	0	0
Si	1	0	0	0	0	8	0
So	0	0	0	3	0	0	0
V	0	0	0	3	0	0	0

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[99]. 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.

---

99 Hoover, D.L. "Multivariate Analysis and the Study of Style Variation" Lit Linguist Computing, (2003) 18, pp.341 - 360.

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

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

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 ek = $(d_{ij} - d_{ij}^{pred})$. The expected value of the squared forecast error is therefore:

---

100   Makarenkov,V. and Legendre, P. "Optimal Variable Weighting for Ultrametric and Additive Trees and K-means Partitioning: Methods and Software". *Journal of Classification* (2001) 18 pp.245-271
101   De Soete,G. "OVWTRE: A program for Optimal Variable Weighting for Ultrametric and Additive Tree Fitting". *Journal of Classification* (1988) 5 pp.101-104

$$E\left(e_k^2\right) = E\left[f\left(\left(x_{ik} - x_{jk}\right)w_k\right) + \varepsilon_k - d_{ij}^{pred}\right]^2$$

This is minimised when

$$f\left(\left(x_{ik} - x_{jk}\right)w_k\right)^2 = d_{ij}^{pred\,2}$$

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 wk. 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.

Change in Agglomeration Coefficient of OVW Dissimilarity Matrix
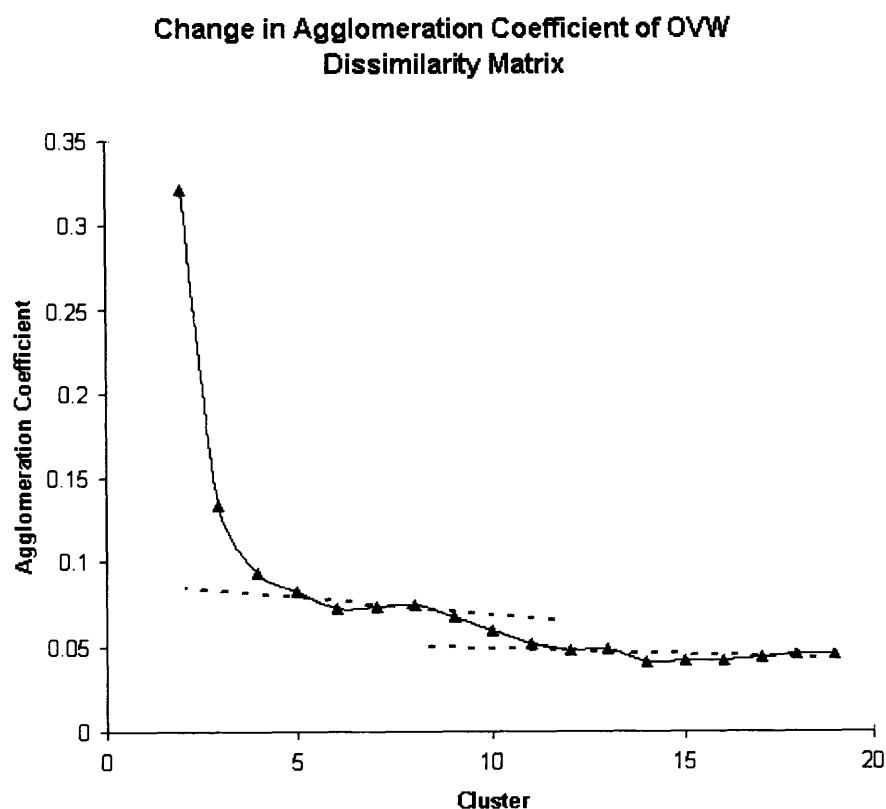
Fig.7.1a *The relative change in the agglomeration coefficient suggests either 5, 6, 9, 10 or 11 clusters.*

Beale's Pseudo-F suggests a 5 cluster solution.

111

Dendrogram by Ward's Method of OVW Data
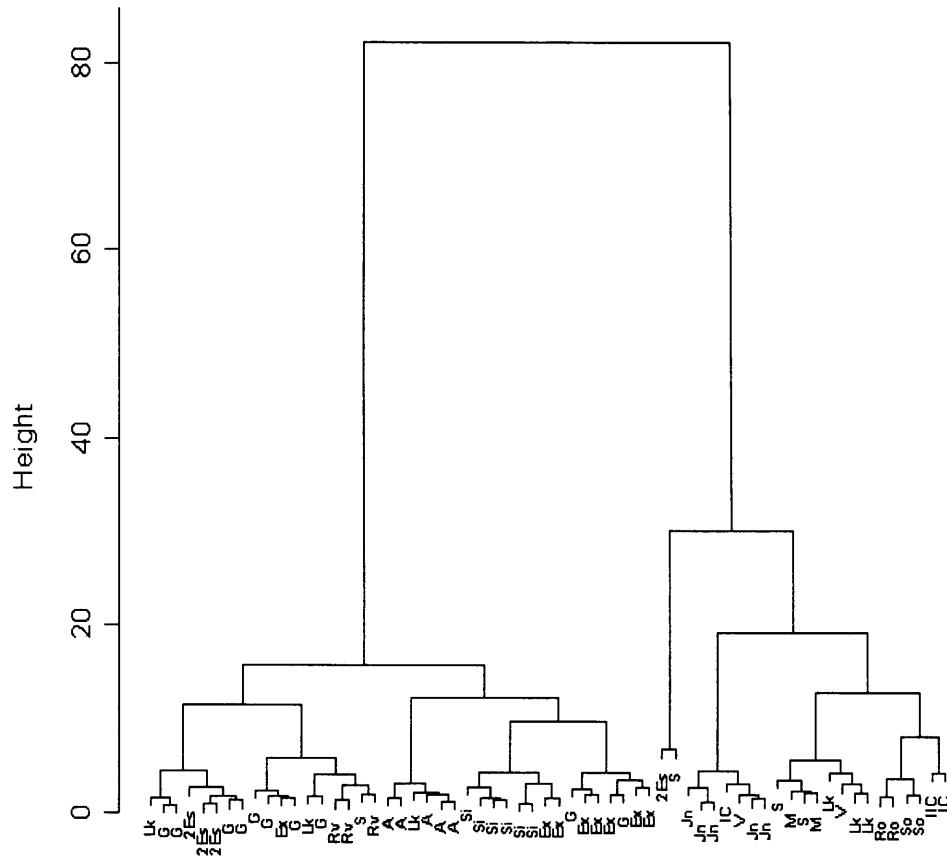
Fig.7.1b Dendrogram obtained from OVW dissimilarity matrix showing a possible 5 clusters.

The results indicate a 5 cluster solution is found, which partitions the cases into the following clusters:

	1	2	3	4	5
2Es	3	1	0	0	0
A	0	0	0	5	0
Ex	1	0	0	7	0
G	8	0	0	2	0
IC	0	0	1	0	1
IIC	0	0	0	0	1
Jn	0	0	5	0	0
Lk	2	0	0	1	3
M	0	0	0	0	2
Ro	0	0	0	0	2
Rv	3	0	0	0	0
S	1	1	0	0	2
Si	0	0	0	6	0
So	0	0	0	0	2
V	0	0	1	0	1

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[103]. Székely and Rizzo[104] 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.
104   Székely, G..J. Rizzo, M.L. "Testing For Equal Distributions In High Dimensions." (2004) obtainable from Department of Mathematics and Statistics, Bowling Green State University, Ohio.

They used the Euclidean distance as their starting point, but define the distance between two distributions, A ={a_1, a_2, ...a_{n1}} and B = {b_1, b_2....b_{n2}} as the e-distance, which, for k=2 is defined by:

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

where $\left\| \ . \ \right\|$ 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
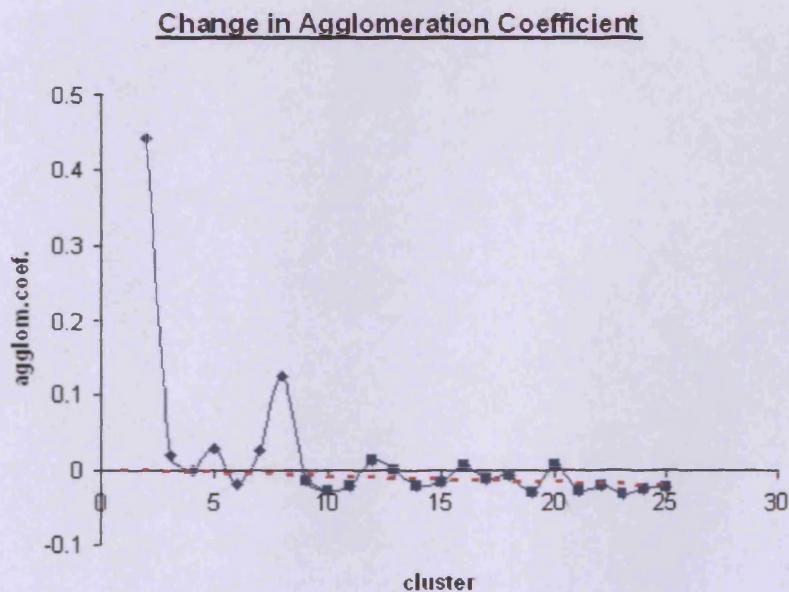
Fig.7.2   Chnge in agglomeration coefficient for Z-transformed
          e-distance clustering. The first significant increase in
          coefficient occurs going from 9 to 8 clusters. A small
          increase is noted going from 6 to 5 clusters. A large increase
          occurs going from 3 to 2 clusters. Trend line shown in red.

Cluster Dendrogram
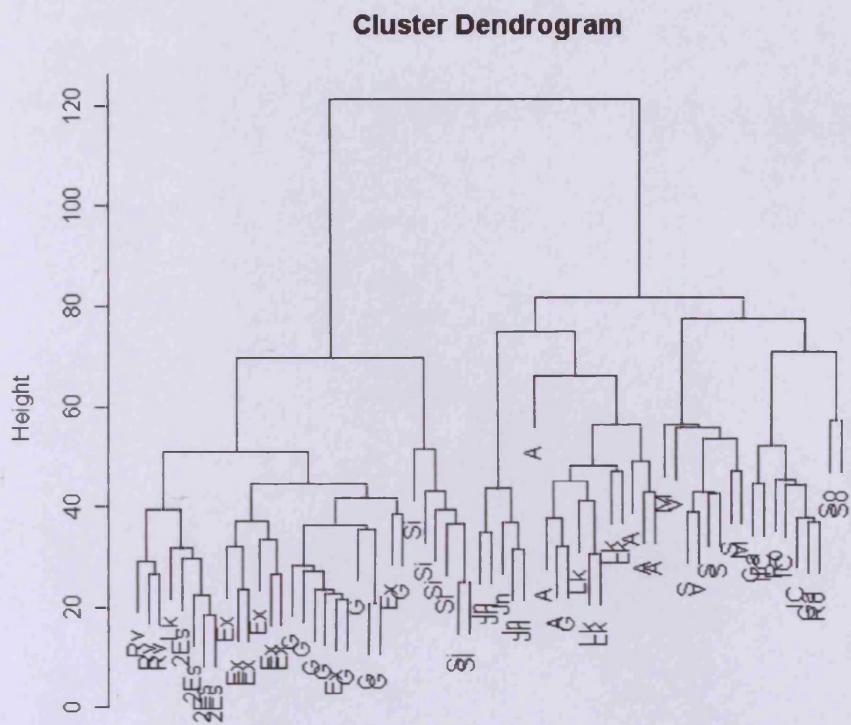
Fig.7.3 Cluster Dendrogram for Z-transformed data using the e-distance method.

Cophenetic Correlation = 0.329

- SQRT-Transformed Data.

Change in Agglomeration Coefficient
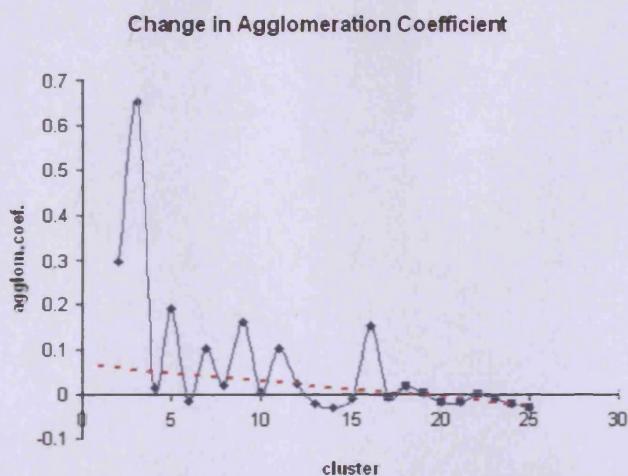
Fig.7.4 Change in agglomeration coefficient for SQRT-transformed data by e-distance. The cluster solution is complicated by the many peaks, but cluster solutions of 17,12,10,8,6,4 are all possibilities. Trend line is shown in red.

Cluster Dendrogram
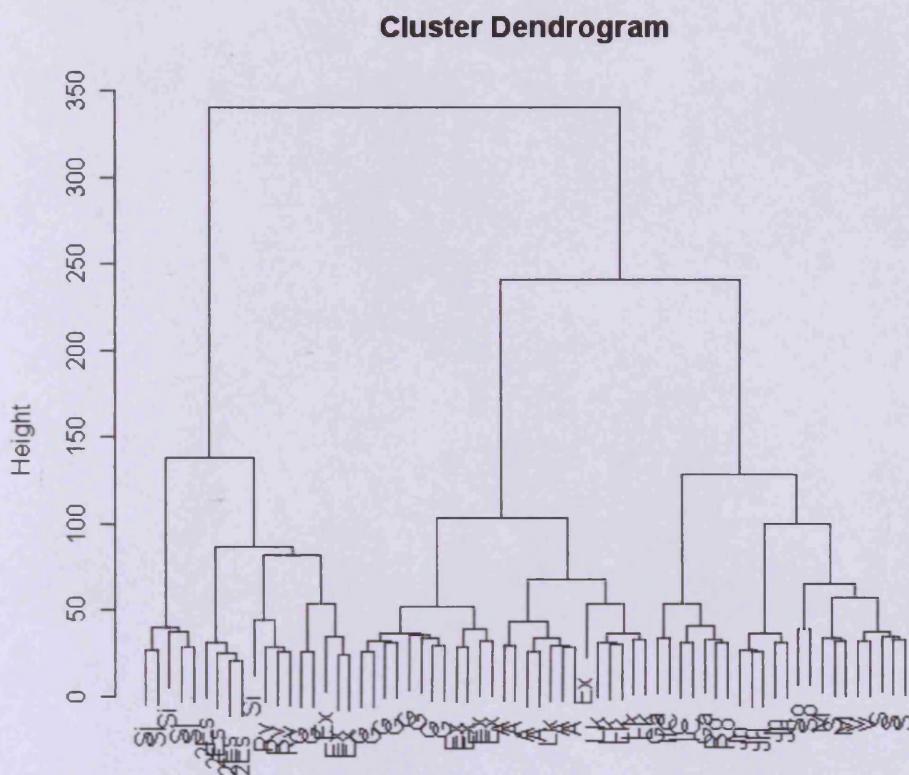
Fig.7.5   *Cluster Dendrogram based on e-distance of SQRT(data)+SQRT(data+1) transformation*

Cophenetic Correlation = 0.685

- RF-Proximity Matrix.

Change in Agglomeration Coefficient
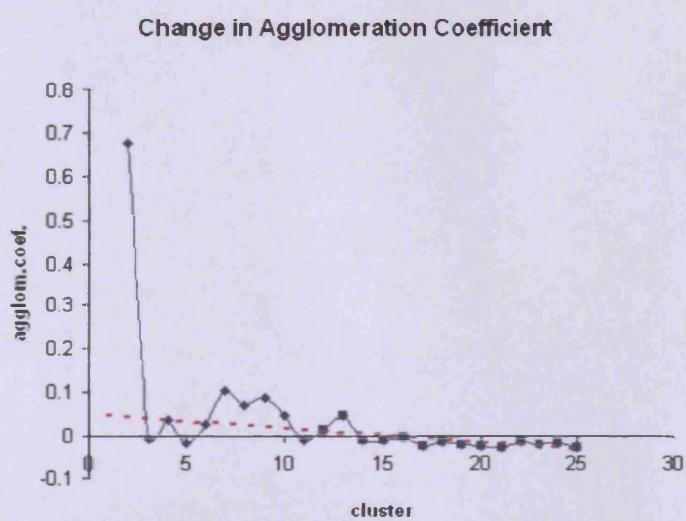
Fig. 7.6   *Change in agglomeration coefficient for RF-proximity matrix by e-distance clustering. Cluster solutions between 13,10,9,&7 are possible.*

116

Cluster Dendrogram
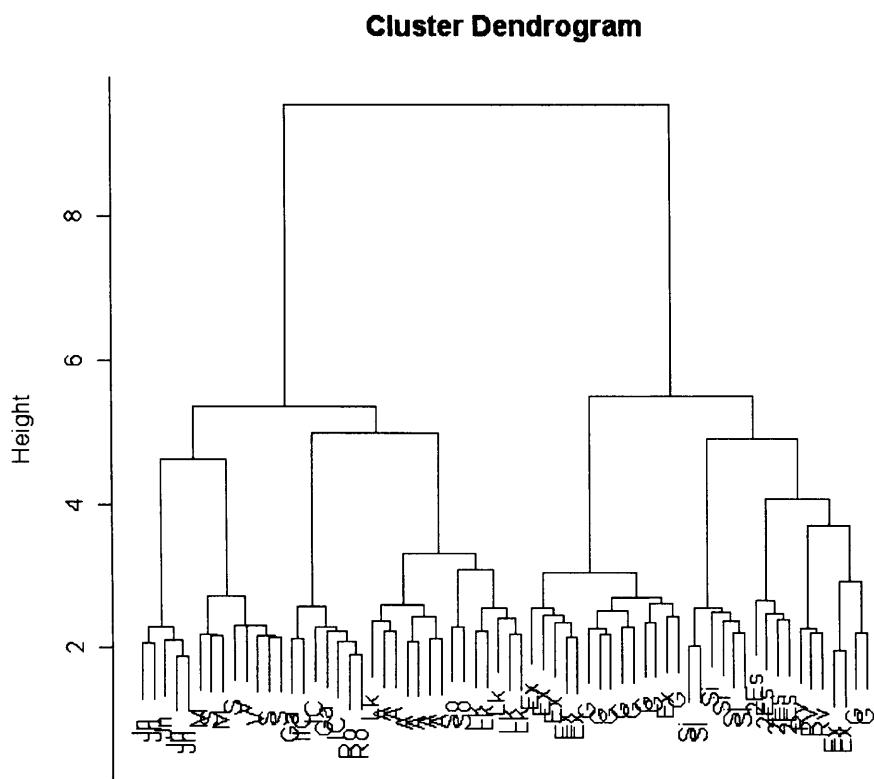
Fig. 7.6 Dendrogram of RF-proximity matrix by e-distance clustering

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.

117

Chapter 8

FINITE MIXTURE DENSITY

Consider the word $\delta\iota\acute{\alpha}$ 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 Histogran an Density Estimate
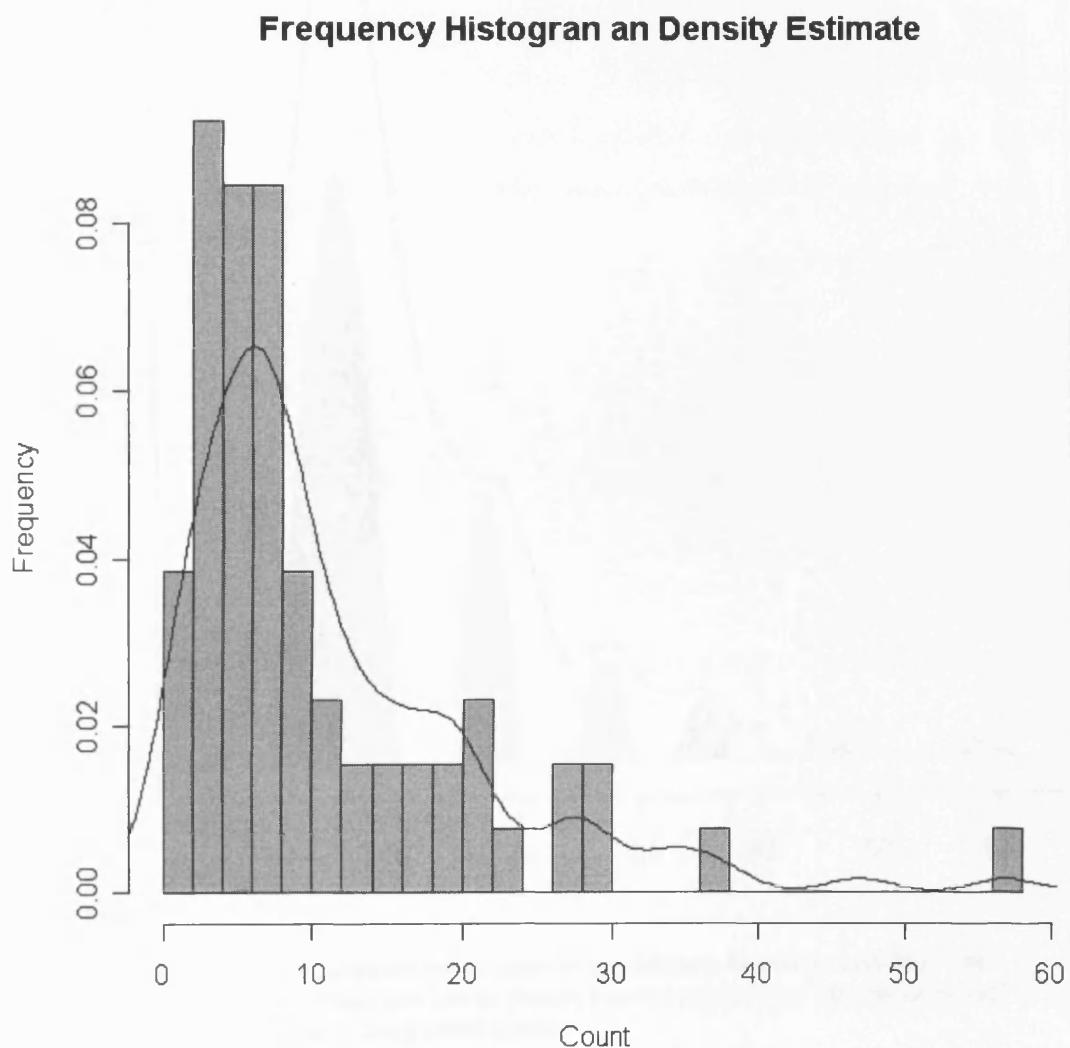
Fig. 8.1 *The frequency histogram and density estimate (black line) using the Sheather-Jones direct plug-in kernel for word-count taken from the 3000word sample block of text. It can be seen that the count is discontinuous and complex. It is not a single 'bell-shaped' curve, but is a skewed density with 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:
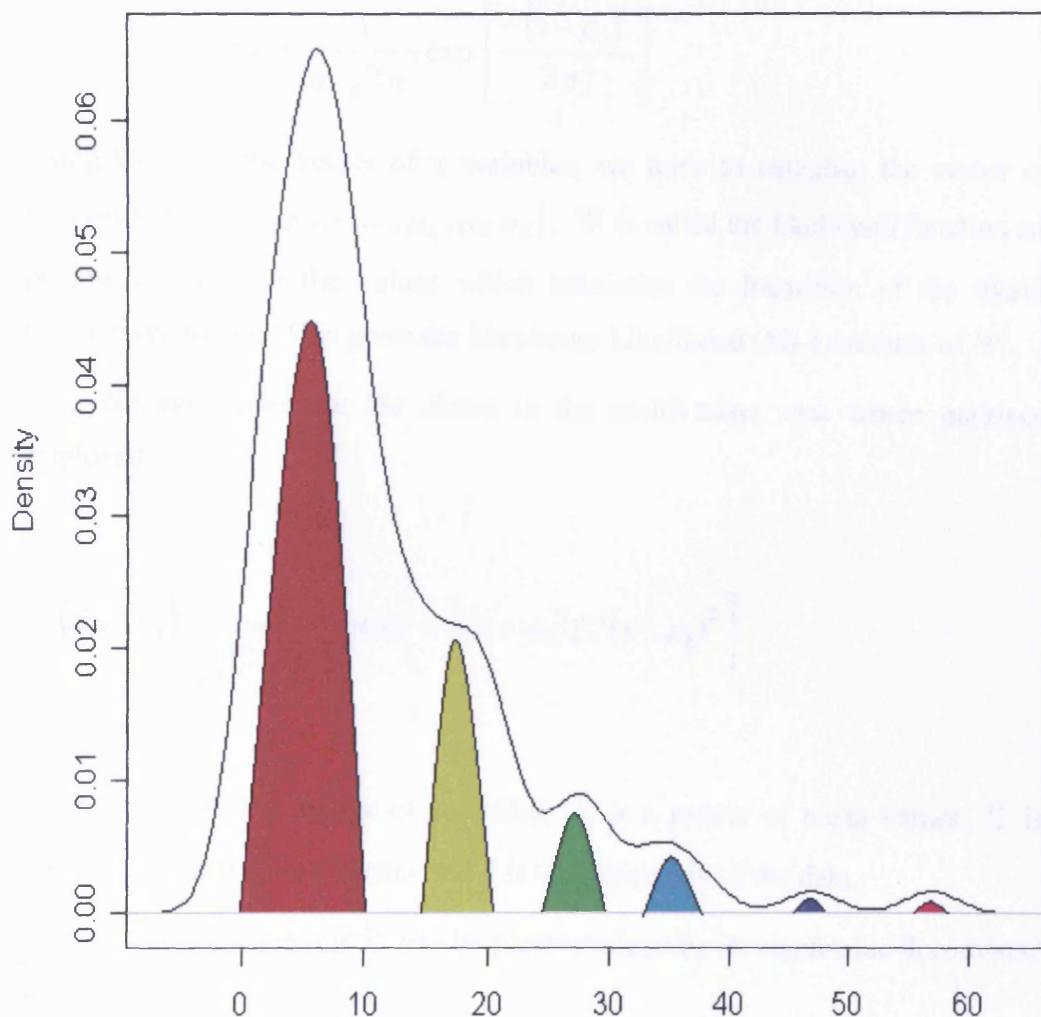
119

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:

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

where π_j is the mixing weight, K is the number of kernels, and $f_j(x)$ is the univariate normal density with mean μ_j and standard deviation σ_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 $\Psi = [\mu_1, \sigma_1, \pi_1, ..., \mu_K, \sigma_K \pi_K]$. Ψ 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 Ψ.

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

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

Where x is now the matrix of variables, μ is a matrix of mean values, Σ 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 λ_k is a scalar multiplier. Fraley and Raftery[105] show that the shape and volume of the cluster

105   Fraley,C. & Raftery,A.E. "MCLUST: Software for Model-Based Clustering, Density Estimation and Discriminant Analysis." Technical Report 415, Department of Statistics: Washington (2003) p.5 http://www.stat.washington.edu/mclust.

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.
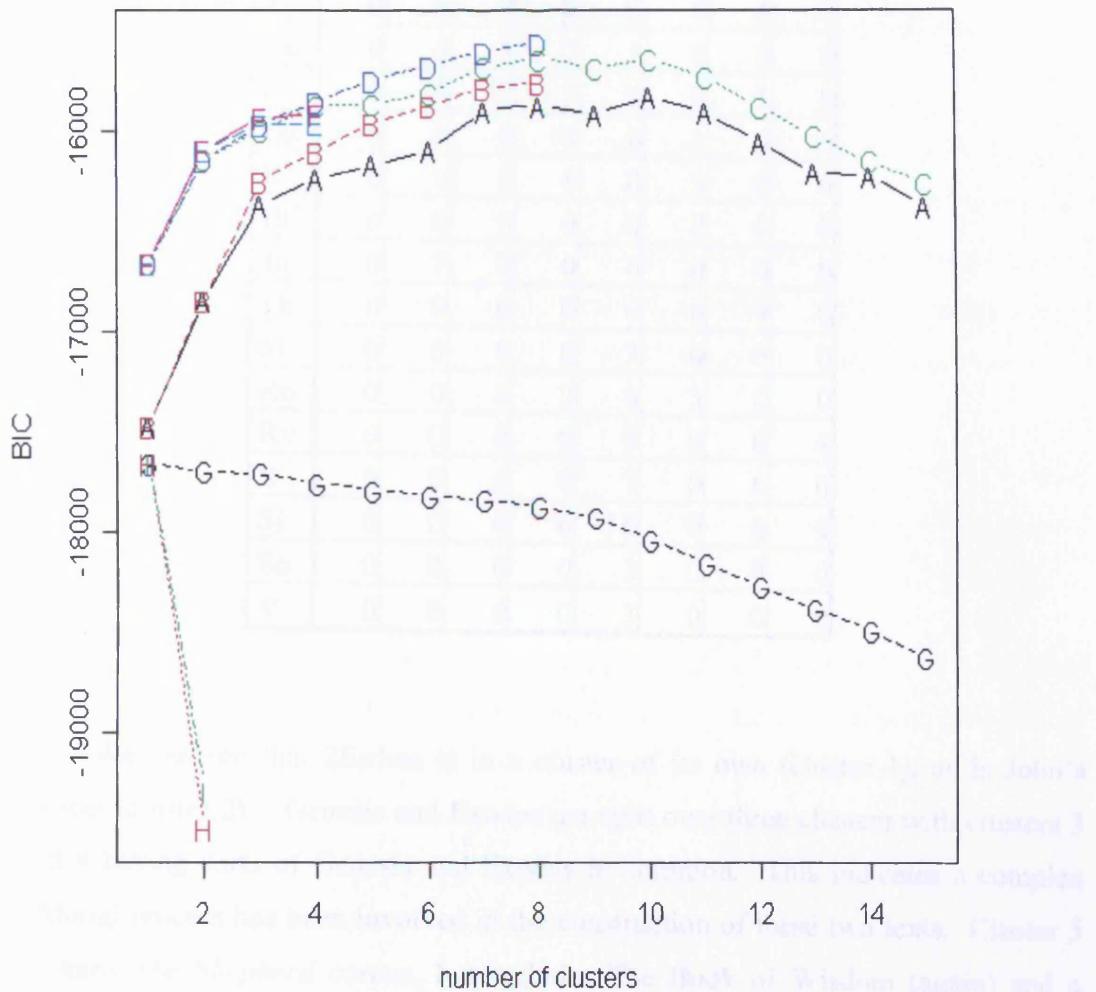
Fig.8.3 *EMclust of SQRT-transformed data using the same data matrix as used for PCA in chapter 6. The best model is 8 clusters type D (diagonal, varying volume, equal shape), although there is not a great deal of difference between this model and models C,B or A.*

BIC (y-axis) is the Bayesian Information Criterion.

	1	2	3	4	5	6	7	8
2Es	6	0	0	0	0	0	0	0
A	0	0	0	0	0	0	8	0
Ex	0	0	4	7	1	0	0	0
G	0	0	2	13	0	0	1	0
Ga	0	0	0	0	0	1	0	0
IC	0	0	0	0	0	3	0	0
IIC	0	0	0	0	0	2	0	0
Jn	0	7	0	0	0	0	0	0
Lk	0	0	0	0	0	0	9	0
M	0	0	0	0	3	0	0	0
Ro	0	0	0	0	0	3	0	0
Rv	0	0	0	0	0	0	0	4
S	0	0	0	0	7	0	0	0
Si	0	0	0	0	0	0	0	9
So	0	0	0	0	3	0	0	0
V	0	0	0	0	3	0	0	0

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:

Block Size	Clusters
3000	14
2500	11
2000	10
1500	14
1000	2
500	2

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   Sample Size = 1000

Cluster	1	2
2Es	26	0
A	3	32
Ex	27	17
G	27	38
IC	0	13
IIC	0	8
Jn	4	27
Lk	6	32
M	2	10
Ro	0	12
Rv	19	0
S	1	28
Si	31	6
So	1	12
V	0	12

Cluster	1	2
2Es	13	0
A	2	15
Ex	21	3
G	26	6
IC	0	6
IIC	0	4
Jn	0	15
Lk	3	16
M	0	6
Ro	0	6
Rv	8	1
S	0	14
Si	5	13
So	0	6
V	0	6

Sample Size = 1500

Cluster	1	2	3	4	5	6	7	8	9	10	11	12	13	14
2Es	8	0	0	0	0	0	0	0	0	0	0	0	0	0
A	0	0	0	0	0	0	0	0	0	11	0	0	0	0
Ex	0	0	4	9	1	1	1	0	0	0	0	0	0	0
G	0	0	0	3	0	2	12	0	0	0	4	0	0	0
IC	0	0	0	0	0	0	0	4	0	0	0	0	0	0
IIC	0	0	0	0	0	0	0	2	0	0	0	0	0	0
Jn	0	10	0	0	0	0	0	0	0	0	0	0	0	0
Lk	0	0	0	0	0	9	0	0	1	2	0	0	0	0
M	0	0	0	0	0	0	0	0	0	0	0	0	4	0
Ro	0	0	0	0	0	0	0	4	0	0	0	0	0	0
Rv	0	0	0	0	0	0	0	0	0	0	0	0	0	6
S	0	0	0	0	0	0	0	0	0	0	0	7	2	0
Si	0	0	0	0	12	0	0	0	0	0	0	0	0	0
So	0	0	0	0	0	0	0	0	4	0	0	0	0	0
V	0	0	0	0	0	0	0	0	0	0	0	4	0	0

Sample Size = 2000

Cluster	1	2	3	4	5	6	7	8	9	10
2Es	6	0	0	0	0	0	0	0	0	0
A	0	0	0	0	0	0	8	0	0	0
Ex	0	0	5	6	1	0	0	0	0	0
G	0	0	3	12	1	0	0	0	0	0
IC	0	0	0	0	0	3	0	0	0	0
IIC	0	0	0	0	0	2	0	0	0	0
Jn	0	7	0	0	0	0	0	0	0	0
Lk	0	0	0	0	7	0	2	0	0	0
M	0	0	0	0	0	0	0	0	3	0
Ro	0	0	0	0	0	3	0	0	0	0
Rv	0	0	0	0	0	0	0	0	0	4
S	0	0	0	0	0	0	0	0	7	0
Si	0	0	0	0	0	0	0	9	0	0
So	0	0	0	0	0	1	2	0	0	0
V	0	0	0	0	0	0	0	0	3	0

Sample Size = 2500

Cluster	1	2	3	4	5	6	7	8	9	10	11
2Es	5	0	0	0	0	0	0	0	0	0	0
A	0	0	0	0	0	0	0	7	0	0	0
Ex	0	0	5	3	1	0	0	0	0	0	0
G	0	0	1	2	9	0	1	0	0	0	0
IC	0	0	0	0	0	2	0	0	0	0	0
IIC	0	0	0	0	0	1	0	0	0	0	0
Jn	0	6	0	0	0	0	0	0	0	0	0
Lk	0	0	0	0	0	0	6	1	0	0	0
M	0	0	0	0	0	0	0	0	0	2	0
Ro	0	0	0	0	0	2	0	0	0	0	0
Rv	0	0	0	0	0	0	0	0	0	0	3
S	0	0	0	0	0	0	0	0	0	5	0
Si	0	0	0	0	0	0	0	0	7	0	0
So	0	0	0	0	0	2	0	0	0	0	0
V	0	0	0	0	0	0	0	0	0	2	0

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
2Es	4	0	0	0	0	0	0	0	0	0	0	0	0	0
A	0	0	0	0	0	0	0	0	5	0	0	0	0	0
Ex	0	0	2	6	0	0	0	0	0	0	0	0	0	0
G	0	0	0	1	0	0	0	0	0	7	2	0	0	0
IC	0	0	0	0	0	0	2	0	0	0	0	0	0	0
IIC	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Jn	0	5	0	0	0	0	0	0	0	0	0	0	0	0
Lk	0	0	0	0	0	6	0	0	0	0	0	0	0	0
M	0	0	0	0	0	0	0	0	0	0	0	0	2	0
Ro	0	0	0	0	0	0	2	0	0	0	0	0	0	0
Rv	0	0	0	0	0	0	0	0	0	0	0	0	0	3
S	0	0	0	0	0	0	0	0	0	0	0	3	1	0
Si	0	0	0	0	0	0	0	6	0	0	0	0	0	0
So	0	0	0	0	2	0	0	0	0	0	0	0	0	0
V	0	0	0	0	0	0	0	0	0	0	0	2	0	0

The increase in the number of clusters in the larger sample sizes resolves Luke-Acts 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:

Cluster	1	2	3	4	5	6	7	8	9	10
2Es	6	0	0	0	0	0	0	0	0	0
A	0	0	0	0	0	0	8	0	0	0
Ex	0	0	5	6	1	0	0	0	0	0
G	0	0	3	12	1	0	0	0	0	0
Ga	0	0	0	0	0	1	0	0	0	0
IC	0	0	0	0	0	3	0	0	0	0
IIC	0	0	0	0	0	2	0	0	0	0
Jn	0	7	0	0	0	0	0	0	0	0
Lk	0	0	0	0	7	0	2	0	0	0
M	0	0	0	0	0	0	0	0	3	0
Ro	0	0	0	0	0	3	0	0	0	0
Rv	0	0	0	0	0	0	0	0	0	4
S	0	0	0	0	0	0	0	0	7	0
Si	0	0	0	0	0	0	0	9	0	0
So	0	0	0	0	0	0	3	0	0	0
V	0	0	0	0	0	0	0	0	3	0

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.

	1	2	3	4	5	6	7	8	9	10	11
2Es	6	0	0	0	0	0	0	0	0	0	0
A	0	0	0	0	0	0	0	0	0	8	0
Ex	0	0	5	4	2	1	0	0	0	0	0
G	0	0	3	13	0	0	0	0	0	0	0
Ga	0	0	0	0	0	0	1	0	0	0	0
IC	0	0	0	0	0	0	3	0	0	0	0
IIC	0	0	0	0	0	0	2	0	0	0	0
Is	0	0	0	0	0	13	0	0	0	0	0
Jn	0	7	0	0	0	0	0	0	0	0	0
Lk	0	0	0	0	6	0	0	0	1	2	0
M	0	0	0	0	0	0	0	0	3	0	0
Ro	0	0	0	0	0	0	3	0	0	0	0
Rv	0	0	0	0	0	0	0	0	0	0	4
S	0	0	0	0	0	0	0	0	7	0	0
Si	0	0	0	0	0	0	0	9	0	0	0
So	0	0	0	0	1	0	1	1	0	0	0
V	0	0	0	0	0	0	0	0	3	0	0

3. Addition of 4 Maccabees:

4 Maccabees was added to balance the Hebrew/Greek text ratio.

	1	2	3	4	5	6	7	8	9	10	11
2Es	6	0	0	0	0	0	0	0	0	0	0
4M	0	0	0	0	0	0	3	0	0	0	0
A	0	0	0	0	8	0	0	0	0	0	0
Ex	0	0	6	5	1	0	0	0	0	0	0
G	0	0	3	13	0	0	0	0	0	0	0
Ga	0	0	0	0	0	0	0	1	0	0	0
IC	0	0	0	0	0	0	0	3	0	0	0
IIC	0	0	0	0	0	0	0	2	0	0	0
Is	0	0	0	0	0	13	0	0	0	0	0
Jn	0	7	0	0	0	0	0	0	0	0	0
Lk	0	0	0	0	9	0	0	0	0	0	0
M	0	0	0	0	0	0	0	0	0	3	0
Ro	0	0	0	0	0	0	0	3	0	0	0
Rv	0	0	0	0	0	0	0	0	0	0	4
S	0	0	0	0	0	0	0	0	0	7	0
Si	0	0	0	0	0	0	0	0	9	0	0
So	0	0	0	0	0	0	3	0	0	0	0
V	0	0	0	0	0	0	0	0	0	3	0

4. Addition of Jeremiah:

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

	1	2	3	4	5	6	7	8	9	10	11	12
2Es	6	0	0	0	0	0	0	0	0	0	0	0
4M	0	0	0	0	0	0	0	0	3	0	0	0
A	0	0	0	0	0	0	0	0	0	0	8	0
Ex	0	0	5	5	1	1	0	0	0	0	0	0
G	0	0	3	12	1	0	0	0	0	0	0	0
Ga	0	0	0	0	0	0	0	0	0	1	0	0
IC	0	0	0	0	0	0	0	0	0	3	0	0
IIC	0	0	0	0	0	0	0	0	0	2	0	0
Is	0	0	0	0	0	12	0	1	0	0	0	0
Jn	0	7	0	0	0	0	0	0	0	0	0	0
Jr	0	0	1	0	0	0	13	0	0	0	0	0
Lk	0	0	0	0	6	0	0	0	0	0	3	0
M	0	0	0	0	0	0	0	0	0	0	0	3
Ro	0	0	0	0	0	0	0	0	0	3	0	0
Rv	0	0	3	0	1	0	0	0	0	0	0	0
S	0	0	0	0	0	0	0	0	0	0	0	7
Si	0	0	0	0	0	0	0	9	0	0	0	0
So	0	0	0	0	0	0	0	0	2	0	1	0
V	0	0	0	0	0	0	0	0	0	0	0	3

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.

	1	2	3	4	5	6	7	8	9	10	11
2Es	6	0	0	0	0	0	0	0	0	0	0
4M	0	0	0	0	1	0	0	2	0	0	0
A	0	0	0	0	0	0	0	8	0	0	0
Ex	0	0	5	5	1	1	0	0	0	0	0
G	0	0	3	13	0	0	0	0	0	0	0
Ga	0	0	0	0	0	0	0	0	1	0	0
IC	0	0	0	0	0	0	0	0	3	0	0
IIC	0	0	0	0	0	0	0	0	2	0	0
Is	0	0	0	0	0	13	0	0	0	0	0
Jn	0	7	0	0	0	0	0	0	0	0	0
Jr	0	0	1	0	0	0	13	0	0	0	0
Lk	0	0	0	0	6	0	0	2	0	0	1
M	0	0	0	0	0	0	0	0	0	0	3
Mt	0	0	0	0	8	0	0	0	0	0	0
Ro	0	0	0	0	0	0	0	0	3	0	0
Rv	0	0	3	0	1	0	0	0	0	0	0
S	0	0	0	0	0	0	0	0	0	0	7
Si	0	0	0	0	0	0	0	0	0	9	0
So	0	0	0	0	1	0	0	1	0	1	0
V	0	0	0	0	0	0	0	0	1	0	2

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[106]. 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:

Similitude contaminated with Revelation	Result
30%	Contaminated Similitude appears in its own group with 1Similitude sample
20%	Contaminated Similitude appears in its own group with 1Similitude sample
10%	Contaminated Similitude appears in its own group with 1Similitude sample
5%	Contaminated Similitude appears in its own group with 1Similitude sample
2%	Contaminated Similitude appears as a singlet in its own group.

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:

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

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:

ἄχρι , ἄχρις ; ἄφνω ; ἄλλως ; ἄν ; ἄνευ ; ἄνω , ἄνωθεν ; ἄρτι ; ἄτερ ; ἀδιαλείπτως ; ἀδίκως ; ἀεί ; ἀφόβως ; ἀφρόνως ; ἀγνῶς ; ἀκριβῶς ; ἀληθῶς ; ἀλλαχόθεν ; ἀλλά , ἀλλ᾽ ; ἀμέμπτως ; ἀναντιρρήτως ; ἀνά ; ἀνδρείως ; ἀνόμως ; ἀντί , ἀνθ᾽ , ἀντ᾽ ; ἀπέναντι ; ἁπλῶς ; ἀποτόμως ; ἀπό , ἀπ᾽ , ἀφ᾽ ; ἀσφαλῶς ; αὔριον ; ἅμα ; ἅπαξ ; χαμαί ; χάριν ; χωρίς ; δεινῶς ; δεῦρο ; δέ ; δή ; διά , δι᾽ ; δικαίως ; διό ; διόπερ ; διότι ; δισσῶς ; δίς ; δυνατῶς ; δυσκόλως ; δωρεάν ; ἑβραϊστί ; ἔμπροσθε , ἔμπροσθεν ; ἔναντι ; ἔνδοθεν ; ἔνθεν ; ἔπειτα ; ἔσω , ἔσωθεν ; ἔτι ; ἔξω , ἔξωθεν ; εἰ , εἴπερ , εἴπως ; ἐάν , ἐάνπερ ; ἐφάπαξ ; ἐγγύς ; ἐκ , ἐξ ; ἐκεῖ , ἐκεῖθεν ; ἐκεῖσε ; ἐκουσίως ; ἐνδόξως ; ἐνθάδε ; ἐνταῦθα ; ἐντεῦθεν ; ἐντός ; ἐνώπιον ; ἐπαύριον ; ἐπάν ; ἐπάνω ; ἐπειδή ; ἐπεί ; ἐπιμελῶς ; ἐπί , ἐπ᾽ , ἐφ᾽ ; ἑτοίμως ; ἐξαίφνης ; ἐξάπινα ; ἑξῆς ; εὐπρεπῶς ; εὐθέως ; εὐθύς ; εὐσεβῶς ; εὐτόνως ; ἕνεκα ; ἕνεκεν ; ἕως ; φανερῶς ; φιλοφρόνως ; φρονίμως ; γάρ ; γέ ; γοῦν ; ἡδέως ; ἡνίκα ; ἤ , ἤτοι ; ἤδη ; ἤ , ἤπερ ; ἱνατί ; ἴσως ; ἰσχυρῶς ; ἵνα , ἵν᾽ ; κἄν ; καί ; καίπερ ; καίτοι ; κἀγώ , κἀμέ ; κἀκεῖ , κἀκεῖθεν ; κακῶς ; καλῶς ; καθά ; καθάπερ ; καθεξῆς ; καθό ; καθότι ; καθώς ; κατά , κατ᾽ , καθ᾽ ; κατέναντι ; κάτω ; κενῶς ; κυκλόθεν ; λαμπρῶς ; λάθρᾳ ; λίαν ; μᾶλλον ; μακρόθεν ; μάλιστα ; μεγάλως ; μενοῦν , μενοῦνγε ; μετά , μετ᾽ , μεθ᾽ ; μεταξύ ; μετρίως ; μέχρι , μέχρις ; μέν ; μέντοι ; μηδαμῶς ; μηδέ , μηδ᾽ ; μηδέποτε ; μή ; μήν ; μήποτε ; μήτε ; μήτι ;

μόλις ; ναί ; νή ; νῦν , νυνί ; ὁμοίως ; ὁμολογουμένως ; ὁμοθυμαδόν ; ὁμοῦ ; ὁσάκις ;
ὁσίως ; ὄπισθε , ὄπισθεν ; ὀπίσω ; ὀργίλως ; ὀρθῶς ; ὀψέ ; οὔπω ; οὔτε ;
οὐ , οὐχ , οὐκ ; οὐχί ; οὐδέ ; οὐδέποτε ; οὐδέπω ; οὐκέτι ; οὕτω , οὕτως ; ὅλως ;
ὅμως ; ὅπου ; ὅπως ; ὅθεν ; ὅταν ; ὅτι ; πανταχοῦ ; παραχρῆμα ; παρά , παρ' ;
παρεκτός ; πάλαι ; πάλιν ; πάντως ; περισσοτέρως ; περί ; πέραν ; πέρυσι , πέρυσιν ;
πικρῶς ; πλήν ; πλουσίως ; πνευματικῶς ; πολλάκις ; ποσάκις ; ποτέ ; πόθεν ;
πόρρω , πόρρωθεν ; πότερον ; πρίν ; προθύμως ; πρό ; πρός ; πρωις , πρωιςθεν ;
πρώτως ; πῶς ; πώποτε ; θαυμαστῶς ; σχεδόν ; σεμνῶς ; σφόδρα ; σήμερον ;
σπουδαίως ; σύν ; ταχέως ; τάχα ; τοιγαροῦν ; τοίνυν ; τότε ; τρίς ; ὑπεράνω ;
ὑπέρ ; ὑποκάτω ; ὑπό , ὑπ' , ὑφ' ; ωςδε ; ὡς ; ὡσαύτως ; ὡσεί ; ὥσπερ ; ὥστε ;

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:

```
>    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
```

Frequency of 100 Best Cluster Determination
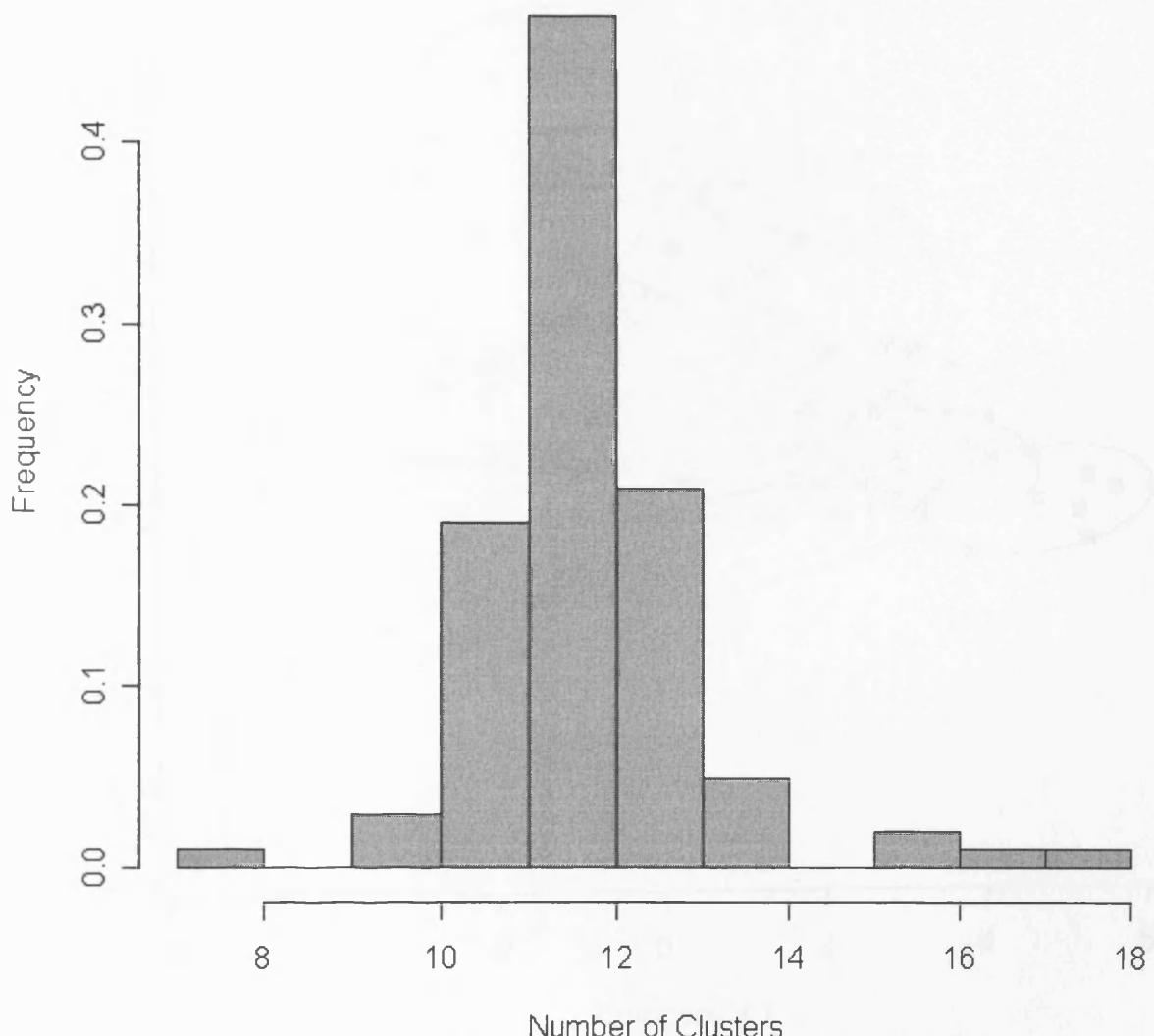


The most frequently determined number of clusters is 11.

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")

PAMClusplot for 11 Clusters
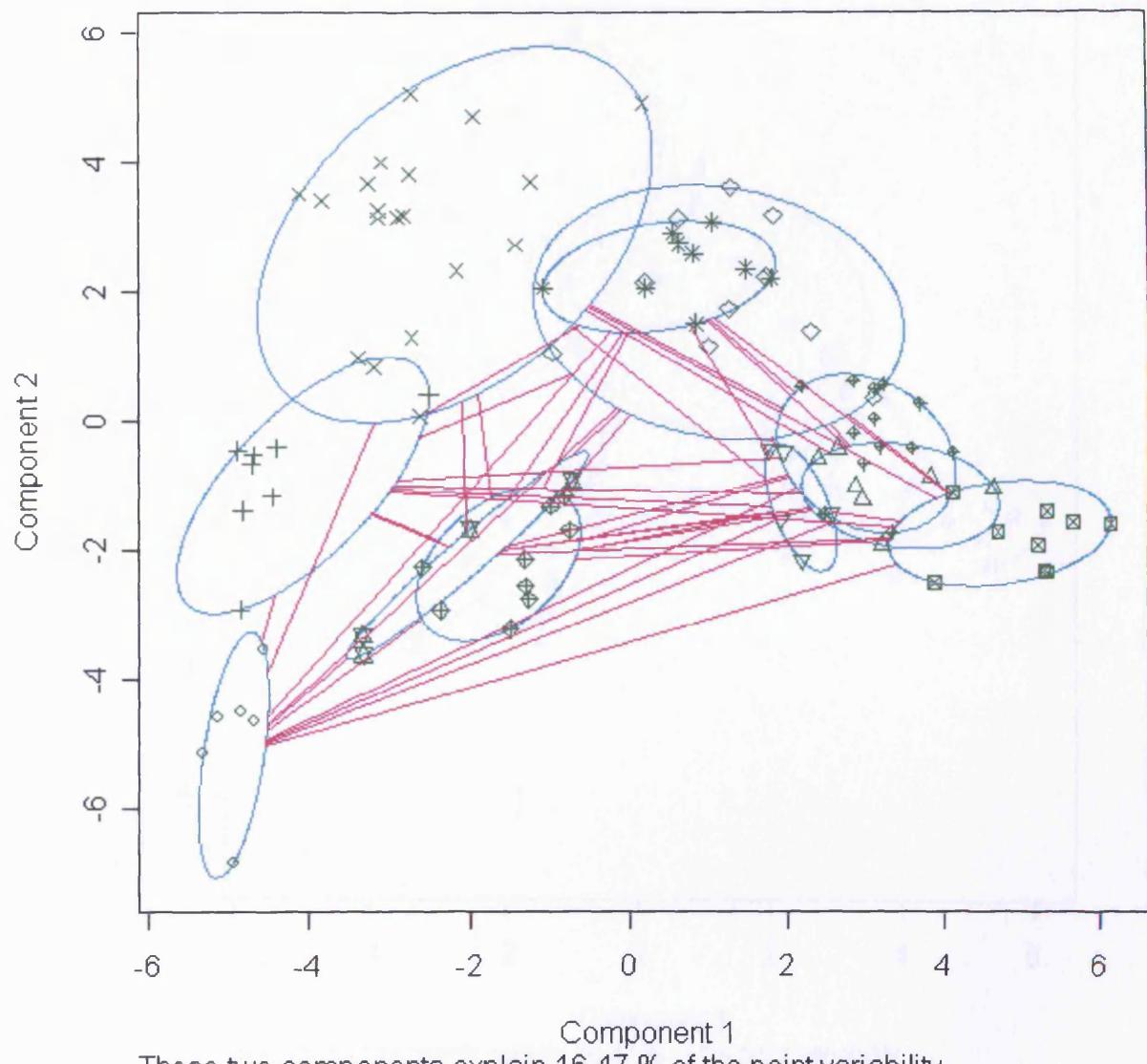
Component 1
These two components explain 16.47 % of the point variability.

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.

clusplot(pam(x = data.urfcombi$proximity, k = 11))
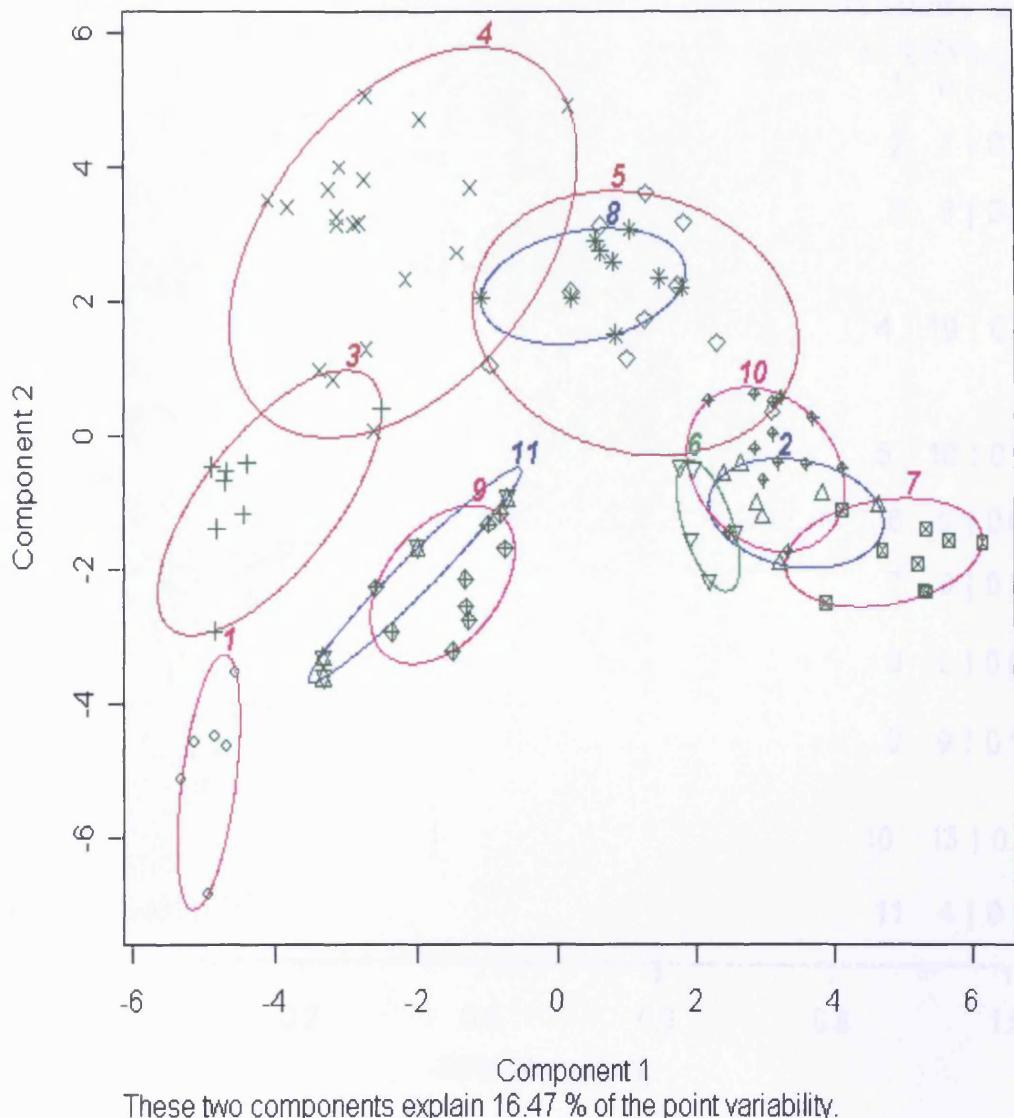
Component 1
These two components explain 16.47 % of the point variability.

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

>   plot(PAMcombi)

Silhouette plot of pam(x = data.urfcombi$proximity, k = 11)

n = 99

11 clusters C_j

$j : n_j \mid ave_{i \in C_j} \ s_i$

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

| | | | | | |
| 0.0 | 0.2 | 0.4 | 0.6 | 0.8 | 1.0 |
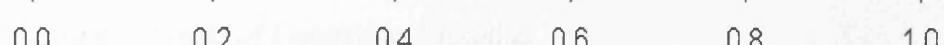
Silhouette width s_i

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:

> 	table(texts, PAMcombi$clustering)

[Results Table 1]

	1	2	3	4	5	6	7	8	9	10	11
2Es	6	0	0	0	0	0	0	0	0	0	0
4M	0	0	0	0	0	3	0	0	0	0	0
A	0	0	0	0	1	0	0	7	0	0	0
Ex	0	0	5	6	1	0	0	0	0	0	0
G	0	0	3	13	0	0	0	0	0	0	0
Ga	0	0	0	0	0	0	1	0	0	0	0
IC	0	0	0	0	0	0	3	0	0	0	0
IIC	0	0	0	0	0	0	2	0	0	0	0
Jn	0	7	0	0	0	0	0	0	0	0	0
Lk	0	0	0	0	7	0	0	2	0	0	0
M	0	0	0	0	0	0	0	0	0	3	0
Ro	0	0	0	0	0	0	3	0	0	0	0
Rv	0	0	0	0	0	0	0	0	0	0	4
S	0	0	0	0	0	0	0	0	0	7	0
Si	0	0	0	0	0	0	0	0	9	0	0
So	0	0	0	0	1	2	0	0	0	0	0
V	0	0	0	0	0	0	0	0	0	3	0

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 5 (questionable integrity) consists of parts of Acts, Exodus, Gospel of Luke, Wisdom of Solomon.

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.

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.

	8	19	20	21
2Es	0	0	0	0
4M	0	0	0	0
A	0	0	0	0
Ex	0	0	0	0
G	0	0	0	0
Ga	1	0	0	0
IC	3	0	0	0
IIC	2	0	0	0
Jn	0	0	0	0
Lk	0	0	0	0
M	0	0	0	3
Ro	3	0	0	0
Rv	0	0	0	0
S	0	1	4	2
Si	0	0	0	0
So	0	0	0	0
V	0	3	0	0

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:

	2Es	4M	A	Ex	G	Ga	IC	IIC	Jn	Lk	M	Ro	Rv	S	Si	So	V	class.error
2Es	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4M	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A	0	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ex	0	0	0	7	5	0	0	0	0	0	0	0	0	0	0	0	0	0.4166667
G	0	0	0	2	14	0	0	0	0	0	0	0	0	0	0	0	0	0.125
Ga	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
IC	0	0	0	0	0	0	2	0	0	0	0	1	0	0	0	0	0	0.3333333
IIC	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	1
Jn	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0
Lk	0	0	0	0	1	0	0	0	0	8	0	0	0	0	0	0	0	0.1111111
M	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	1
Ro	0	0	0	0	0	1	1	0	0	0	0	1	0	0	0	0	0	0.6666667
Rv	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0
S	0	0	0	0	0	0	0	0	0	0	1	0	0	6	0	0	0	0.1428571
Si	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0
So	0	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
V	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	1

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]

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2Es	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4M	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0
A	0	0	8	0	0	0	0	0	0	0	0	0	0	0	0
Ex	0	0	0	7	5	0	0	0	0	0	0	0	0	0	0
G	0	0	0	2	14	0	0	0	0	0	0	0	0	0	0
Ga	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
IC	0	0	0	0	0	0	2	0	0	0	0	1	0	0	0
IIC	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0
Jn	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0
Lk	0	0	0	0	1	0	0	0	0	8	0	0	0	0	0
M	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0
Ro	0	0	0	0	0	1	1	0	0	0	0	1	0	0	0
Rv	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0
S	0	0	0	0	0	0	0	0	0	0	1	0	0	6	0
Si	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9
So	0	1	1	0	0	0	0	0	0	0	0	0	0	0	1
V	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0

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 consist 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
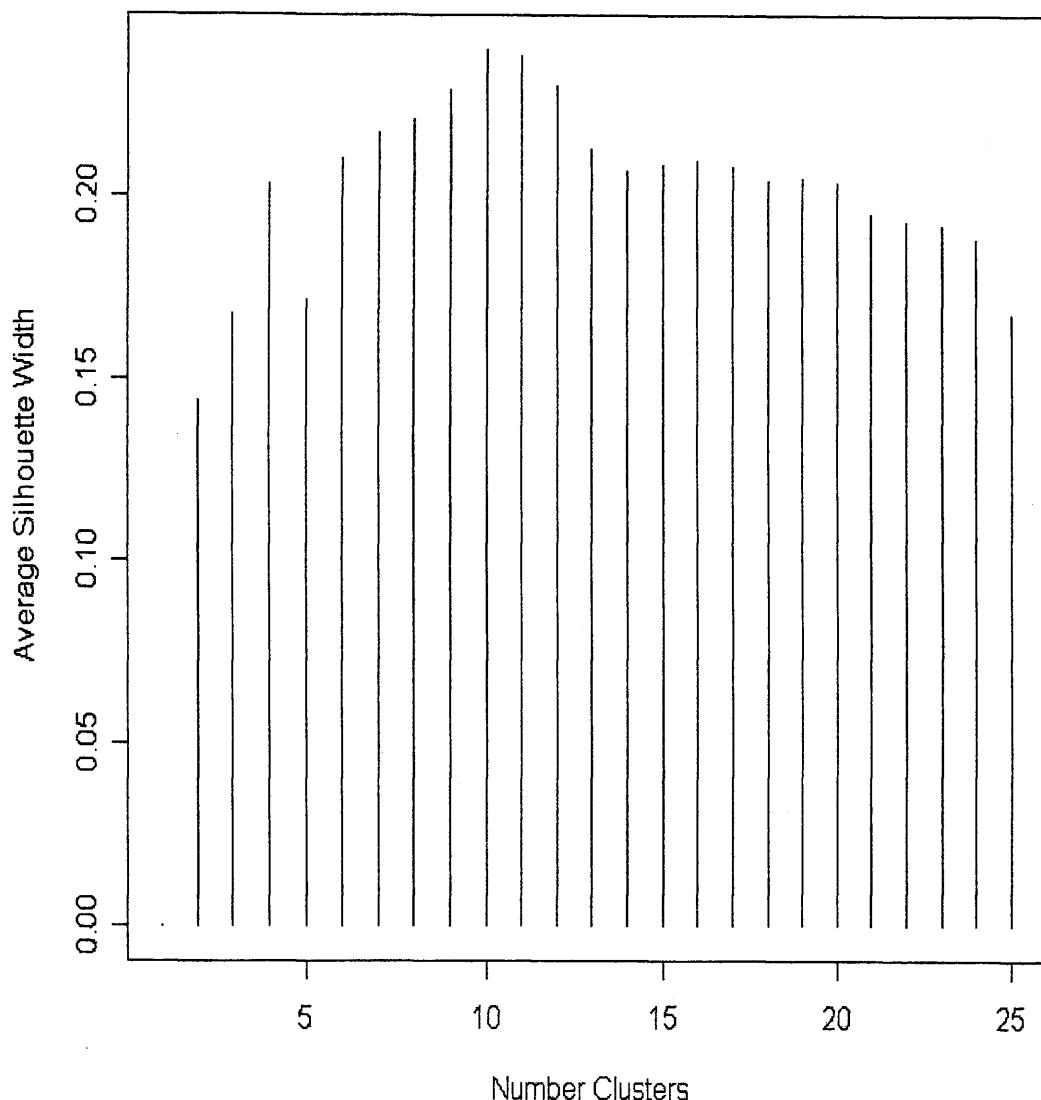
A silhouette plot for the 10 cluster solution is:

Silhouette plot of pam(x = data.rf$proximity, k = 10)

n = 99

10 clusters C_j
$j: n_j \mid ave_{i \in C_j} \, s_i$
1 : 6 | 0.51
2 : 7 | 0.43
3 : 14 | 0.14
4 : 15 | 0.08
5 : 9 | 0.18
6 : 9 | 0.37
7 : 8 | 0.30
8 : 14 | 0.31
9 : 13 | 0.14
10 : 4 | 0.26

Silhouette width s_i
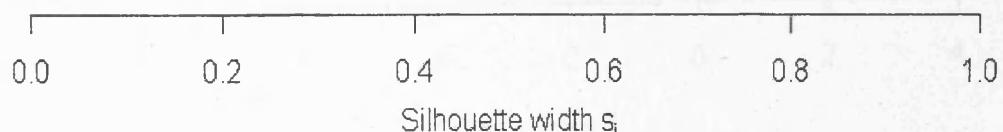
Average silhouette width : 0.24

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
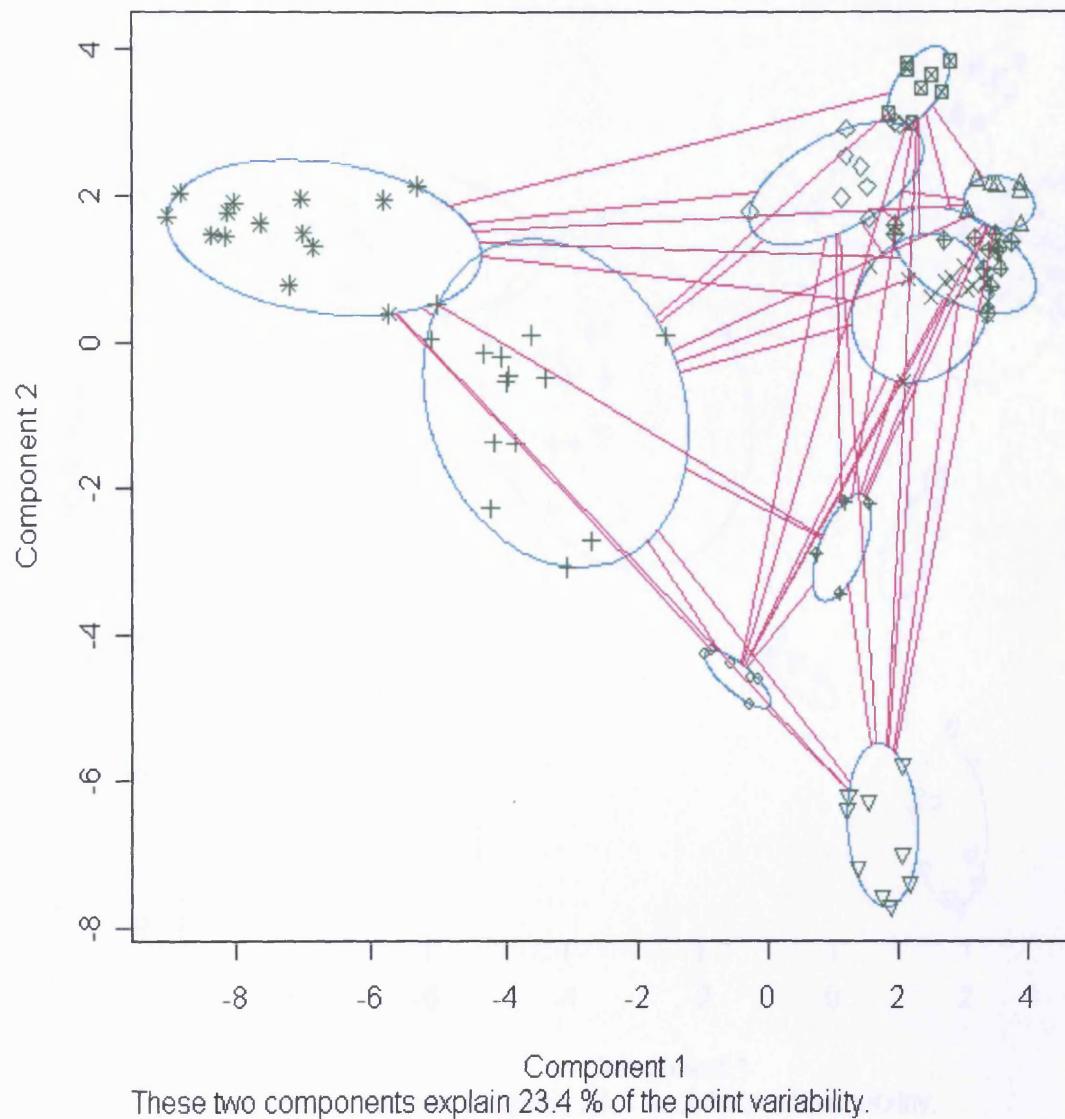
Component 1
These two components explain 23.4 % of the point variability.

Supervised Random Forest with 10 Clusters
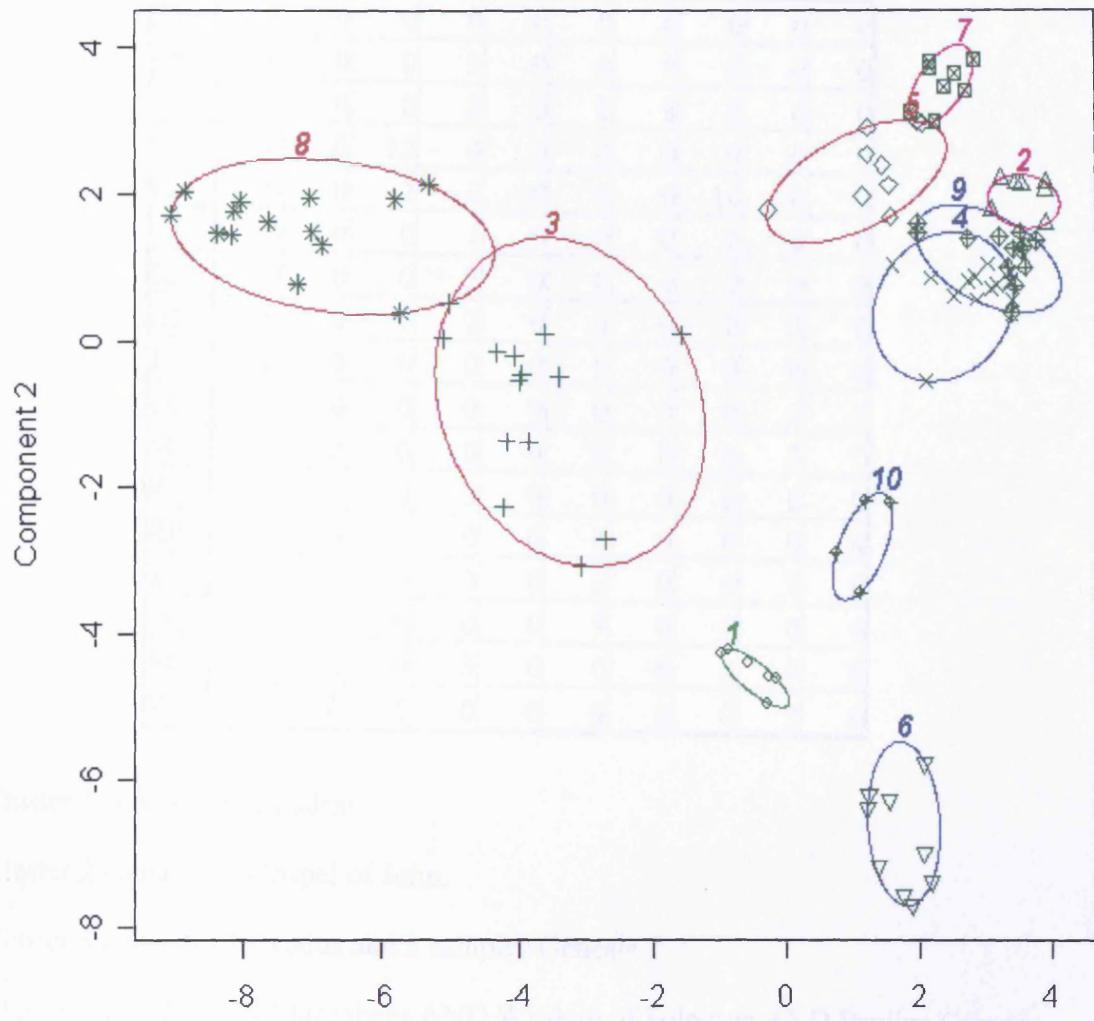
Component 1
These two components explain 23.4 % of the point variability.

[Results Table 3]

	1	2	3	4	5	6	7	8	9	10
2Es	6	0	0	0	0	0	0	0	0	0
4M	0	0	0	3	0	0	0	0	0	0
A	0	0	0	0	0	0	8	0	0	0
Ex	0	0	12	0	0	0	0	0	0	0
G	0	0	2	0	0	0	0	14	0	0
Ga	0	0	0	1	0	0	0	0	0	0
IC	0	0	0	3	0	0	0	0	0	0
IIC	0	0	0	2	0	0	0	0	0	0
Jn	0	7	0	0	0	0	0	0	0	0
Lk	0	0	0	0	9	0	0	0	0	0
M	0	0	0	0	0	0	0	0	3	0
Ro	0	0	0	3	0	0	0	0	0	0
Rv	0	0	0	0	0	0	0	0	0	4
S	0	0	0	0	0	0	0	0	7	0
Si	0	0	0	0	0	9	0	0	0	0
So	0	0	0	3	0	0	0	0	0	0
V	0	0	0	0	0	0	0	0	3	0

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 5 consists of Gospel of Luke.

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

method , although it confirms the single author hypothesis.

Finite Mixture Density.

The number of variables must be less than the number of cases. Multicolinearity 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 multicolinearity. 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.

> 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:

ἵνα , ἵν' ;   ; καί ; μετά ,   μετ' ,   μεθ' ; μέν ; μή ; οὐ ,   οὐχ ,   οὐκ ; ὅταν ; ὅτι ; ἀλλά ,   ἀλλ' ; πρό' ; πρωι' ,   πρωι'θεν ; σύν ; ὑπό ,   ὑπ' ,   ὑφ' ; ὥ' ; ἀπό ,   ἀπ' ,   ἀφ' ; δέ ; διά ,   δι' ; ἔναντι ; ἐκ ,   ἐξ ; ἐκεῖ ,   ἐκεῖθεν ; ἐπί ,   ἐπ' ,   ἐφ' ; ἕω' ; γάρ ; ἡνίκα ;

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

> 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.

151
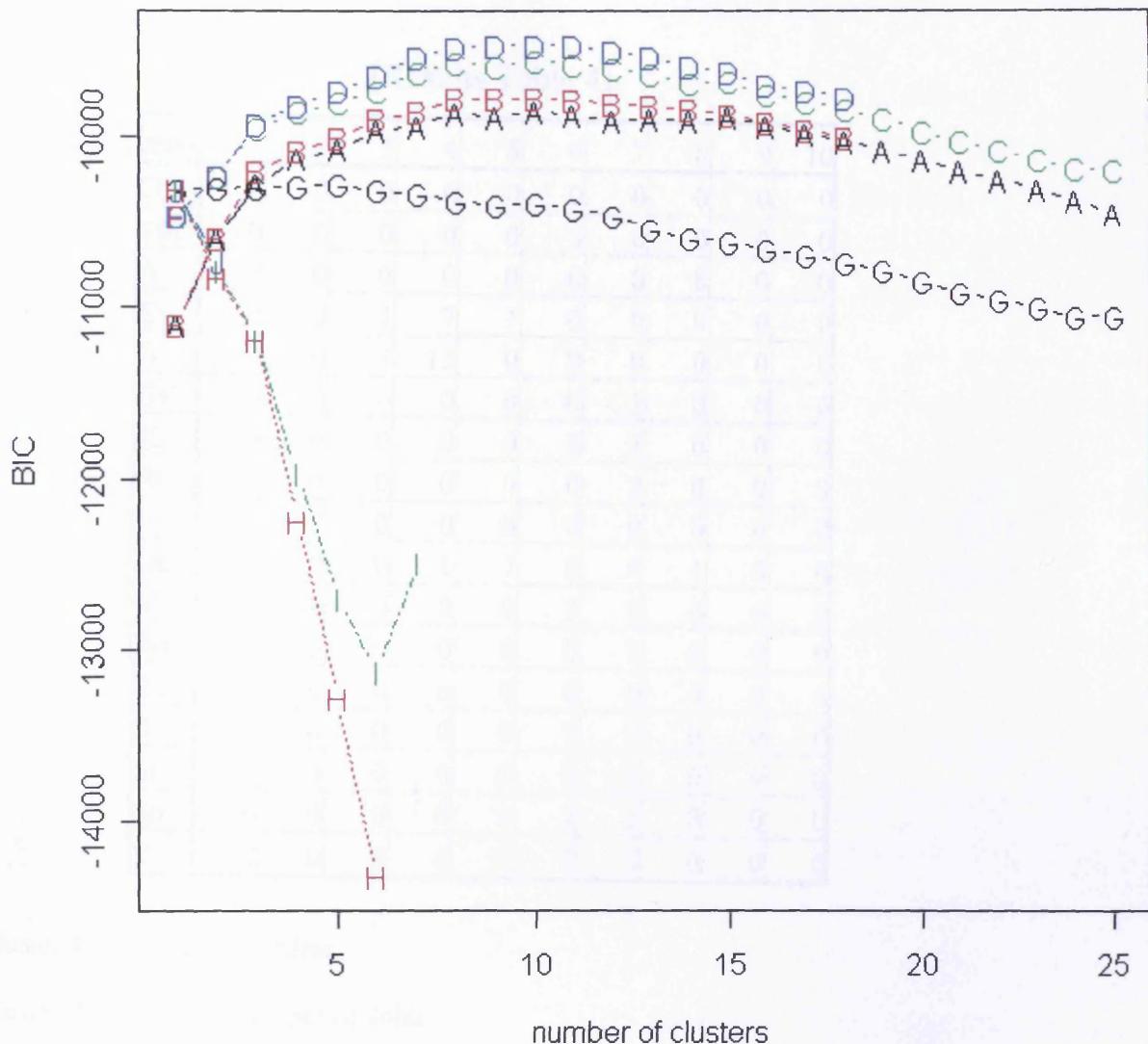
EII VII EEI VEI EVI VVI EEE EEV VEV VVV

"A" "B" "C" "D" "E" "F" "G" "H" "I" "J"

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:

>   data.em3 <- summary(data.em, SQRTdata, G=10, modelNames="VEI")

>   table10 <- table(texts,data.em3$classification)

The following table of classification was obtained:

[Results Table 4]

texts	1	2	3	4	5	6	7	8	9	10
2Es	6	0	0	0	0	0	0	0	0	0
4M	0	0	0	0	0	3	0	0	0	0
A	0	0	0	0	0	0	0	8	0	0
Ex	0	0	4	7	1	0	0	0	0	0
G	0	0	3	13	0	0	0	0	0	0
Ga	0	0	0	0	0	0	1	0	0	0
IC	0	0	0	0	0	0	3	0	0	0
IIC	0	0	0	0	0	0	2	0	0	0
Jn	0	7	0	0	0	0	0	0	0	0
Lk	0	0	0	1	7	0	0	1	0	0
M	0	0	0	0	0	3	0	0	0	0
Ro	0	0	0	0	0	0	3	0	0	0
Rv	0	0	0	0	0	0	0	0	0	4
S	0	0	0	0	0	7	0	0	0	0
Si	0	0	0	0	0	0	0	0	9	0
So	0	0	0	0	0	2	1	0	0	0
V	0	0	0	0	0	3	0	0	0	0

Cluster 1 consists of 2 Esdras.

Cluster 2 consists of Gospel of John.

Cluster 3 consists of Exodus and Genesis.

Cluster 4 consists of Genesis, Exodus AND part of Gospel of Luke.

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 ma be the Lukan corpus) consists of the majority of Acts and two cases from Luke. These are interesting connections and deserve more attention[108] 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:
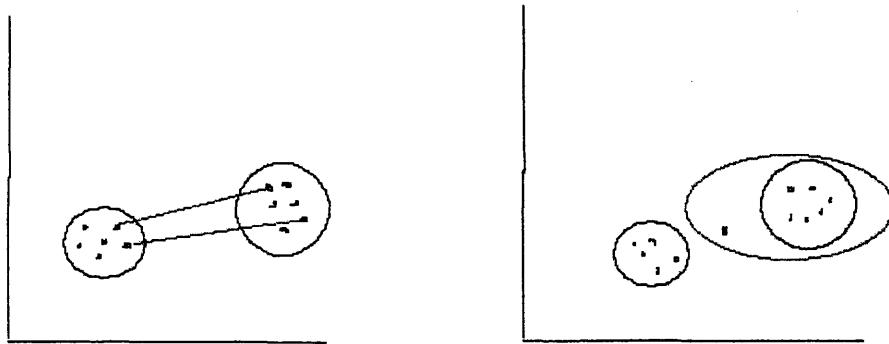
In the illustration on the left, two clusters are well separated. In the illustration on the right, the one chaining sample combines with a cluster to form a new cluster (red ellipse). This new cluster is now closer to the other cluster than would be expected, simply because of the one data point. Such chaining would explain why the problems arise with Luke and parts of Genesis and Exodus[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

---

109 The Septuagintal style of Luke is noted and discussed in Fitzmeyer *The Gospel According to Luke*. Chaining would therefore occur between those sections of Luke containing most Septuagintal features ie. Infancy stories.

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 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 seem 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[110] 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

---

110 Labov, W. Waletzky, J. "Narrative Analysis: Oral Versions of Personal Experience." In Helm, J. (ed) *Essays on the Verbal and Visual Arts.* pp.12-44

going to Cumae". Although the original text is unclear, a choice between εἰς κώμας (to the countryside), and εἰς 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" (Vl.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" (Vl.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?" (Vl.2.2).

The narrative now takes a recursion, defining a new orientation - "While I was considering and doubting these things in my heart" (Vl.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" (V1.3.1). 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" (Vl.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" (Vl.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[111] 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,

---

111 Ryave, A. "On the achievement of a series of stories." in Schenkein, J. (ed.) *Studies in the Organization of Conversation.* pp.113-132

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:

READER ⟸ HERMAS ⟸⟹ VISION CHARACTER ⟸ WRITER
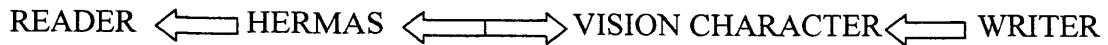
The quest for understanding is played out in the interaction between Hermas and Vision Character and it proceeds via a question and answer format (V1.1.7&8). The question and answer format is basic to our understanding of the whole work of *The Shepherd.* The writer informs and influences his reader 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:

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

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 ($\pi\rho\sigma\epsilon\acute{u}\chi\sigma\upsilon$ - 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} <u>A progressive argument making use of the adversative conjunction.</u>

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 $\dot{\alpha}\lambda\lambda\dot{\alpha}$, '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!).

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

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
113 Huddleston, R., Hudson, R.A., Winter, E., Henrici, A. *Sentence and Clause in Scientific English.* p.125ff

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.*

'Αλλ'
|
οὐχ ἔνεκα τούτου
|
ὁ θεός ὀργίζεται σοι
|
ἀλλ'
|
ἵνα (you) ἐπιστρέψῃς τόν οἶκόν
|
σου
|
τὸν ἀνομήσαντα
|
εἰς τόν κύριον
⋮
καὶ
⋮
εἰς ὑμᾶς
|
τοὺς γονεῖς
|
αὐτῶν

167

ἀλλ᾽

ἡ πολυσπλαγχνία ἠλέησέν σε

τοῦ κυρίου

καὶ

καὶ τόν οἶκον σου

ἰσκυροποιήσει

καί σε

θεμελιώσει ἐν τῇ δόξῃ αὐτοῦ

μόνον

μὴ ῥαθυμήσῃς

ἀλλὰ

εὐψύξει

καὶ τόν οἶκον σου

ἰσχυροποίει
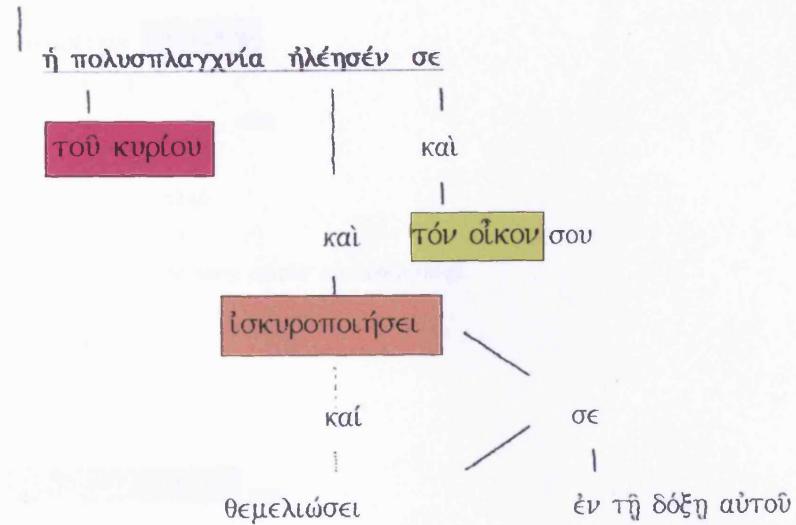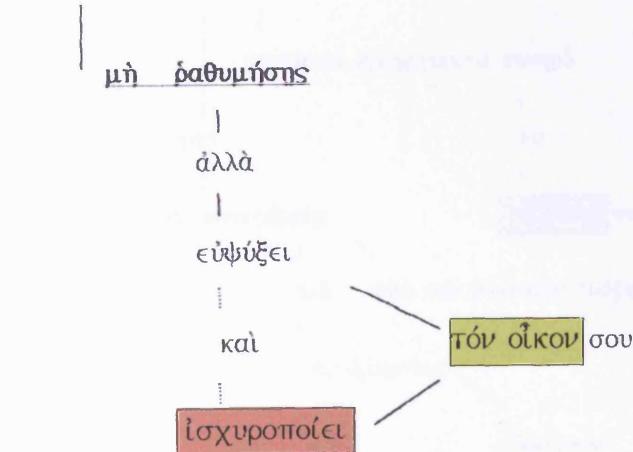
ἀλλὰ
|
ὢν φιλότεκνος
|
(you) ἐνουθέτεις τὸν οἶκον
                    | |
                   οὐκ   σου
                       \
                      ἀλλὰ
                       |
                  ἀφῆκες αὐτὸν καταφθαρῆναι

διὰ τοῦτό
          | ὁ κύριος ὀργίζεται σοι
          |
          |
          ἀλλά
          |
          ἰάσεται σου
          |
          πάντα τά προγεγονότα πονηρὰ
          | |
         γαρ   ἐν
          | |
      σὺ κατεφθάρης   τῷ οἶκῳ σου
          | \
         διὰ   ἀπὸ τῶν βιωτικῶν πράξεων
          |
        τὰς ἁμαρτίας
                    \
           καὶ   ἐκείνων
                         \
          ἀνομήματα

169

ἀλλ᾽

ἡ πολυσπλαγχνία ἠλέησέν σε

τοῦ κυρίου   καὶ

καὶ   τόν οἶκον σου

ἰσκυροποιήσει

καί   σε

θεμελιώσει   ἐν τῇ δόξῃ αὐτοῦ

μόνον

μὴ ῥαθυμήσῃς

ἀλλὰ

εὐψύξει

καὶ   τόν οἶκον σου

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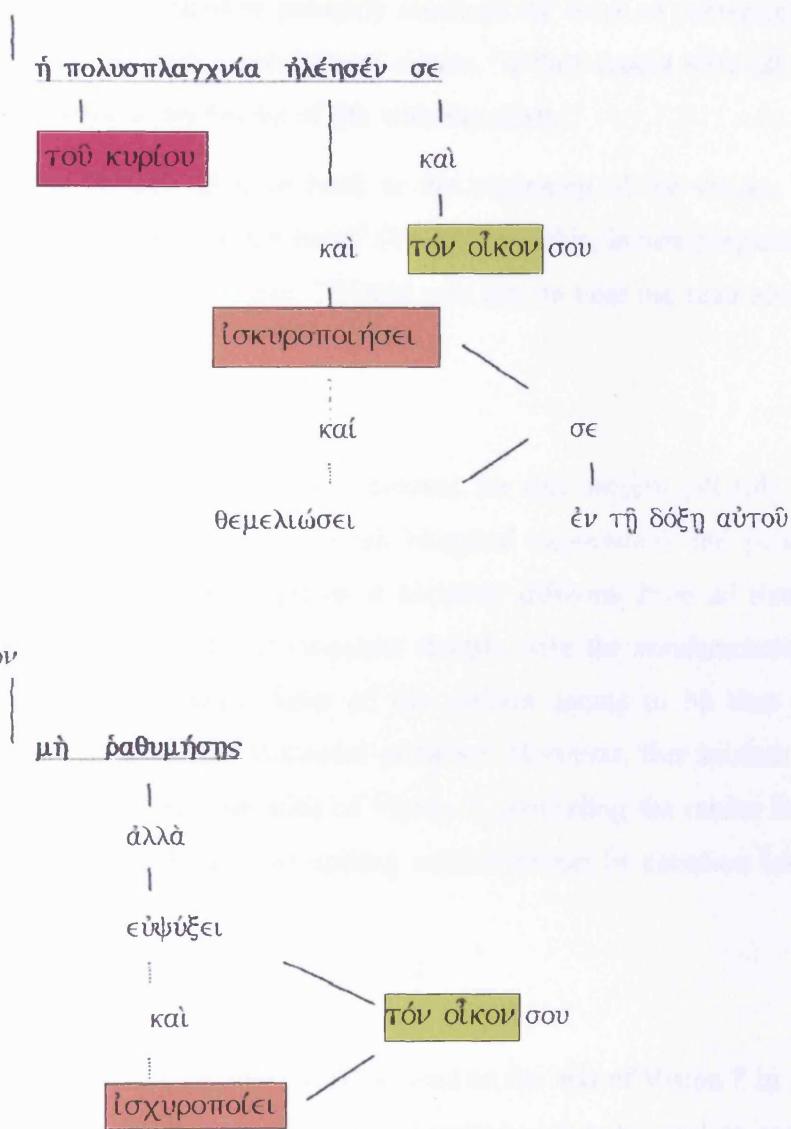
Not only can we see the function of the ἀλλά phrases, but we now see a number of interesting structural signals become apparent (note the highlighted words). We will come back to these later in this chapter.

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[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

---

115 Osiek,C. *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 (V1.1.3), the Visionary Character, Rhoda, describes God as the one who created (V1.1.6), and the Visionary Character, the old woman, reads a description of God the creator (V1.3.4). 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.
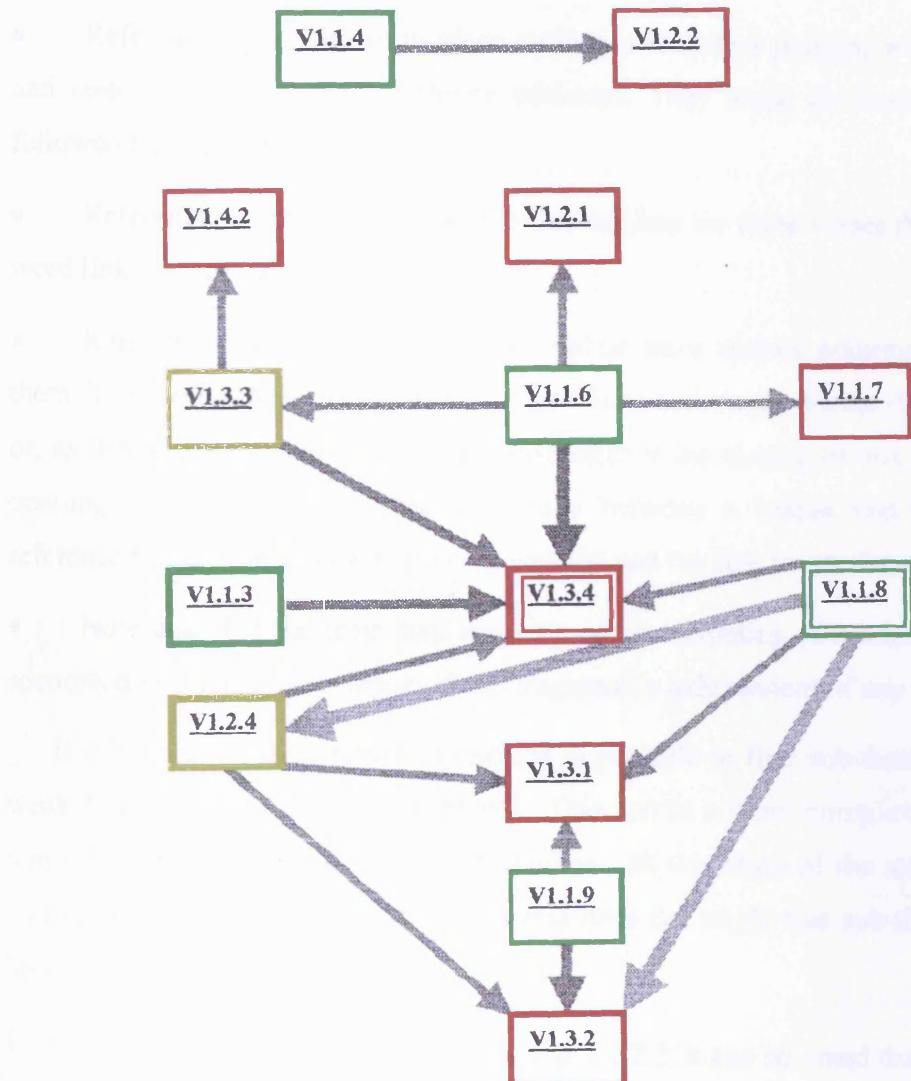
The width and colour of the arrows is proportional to the bond strength.

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.

- References 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)*[118] the female Visionary Character who greets him *(ἀσπάζομαι)* using the same greetings formula *('Ερμᾶ, χαῖρε)*.

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

---

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 Vl.1.8 and Vl.2.4, and between Vl.1.8 and Vl.3.2. Vl.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. Vl.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.
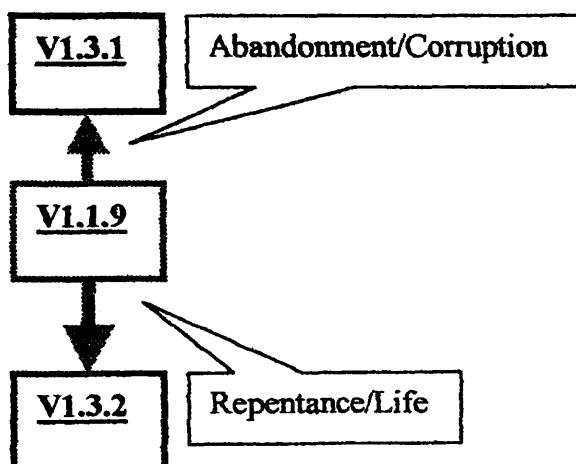
These contrasting outcomes form a conclusion to themes in Vl.1.8 and Vl.2.4. The bottom part of the 'map' displays the structure of theme development in the narrative.

V1.2.4   V1.3.1   V1.1.8

V1.1.9

V1.3.2
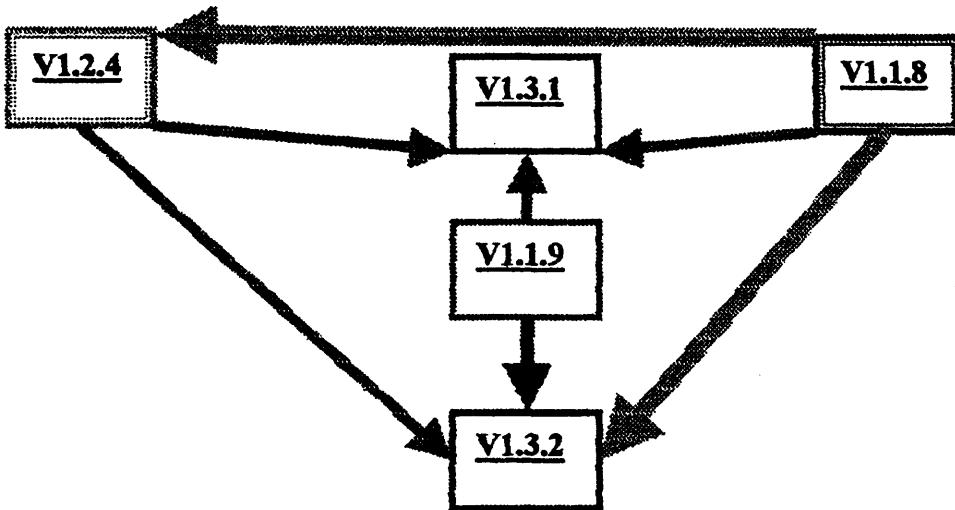
V1.2.4 is a restatement by the second Visionary Character of the discourse uttered by the first Visionary Character. There is a slight difference in emphasis between the two statements. In VI.I.8, the wicked desire (ἡ πονηρά ἐπιθυμια) is regarded as an evil deed (πονηρόν πρᾶγμα), and both the desire 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. Vl.2.4 changes the πονηρόν πρᾶγμα of Vl.1.8 to a πονηρόν ἔργον 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 πρός τό ἐπιθυμῆσαι αὐτήν
ἤδη ἐμοίχευσεν αὐτήν ἐν τῇ' καρδίᾳ αὐτοῦ..

Hermas' ἐπιθυμία is described as an evil πρα'γμα (Vl.1.8) and as an evil ἔργον (Yl.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 μηνπεφρικώς καί λυπούμενος (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:
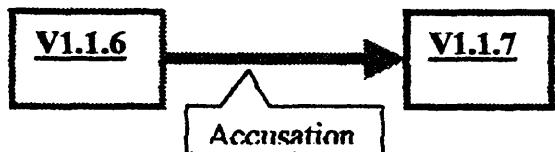

V1.1.6 and V1.1.7 suggest a courtroom scene and deals with the accusation and the cross-examination. A concerned Hermas asks Rhoda (the Vision Character), "Are you now accusing me?" ie. setting out the **words** (implied) of accusation. "No," she said, "but listen to the **words** which I am going to 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:
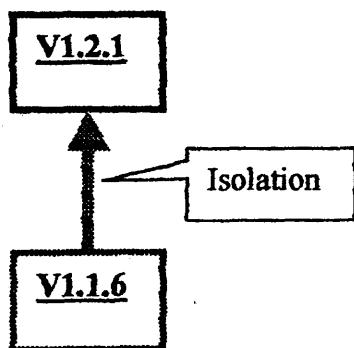
V1.1.6 and V1.2.1 display Hermas' isolation. V1.2.1 begins, "After she had spoken these **words** the **heavens** were shut...". This does not simply mark the closing of the vision. The Visionary Character has reminded Hermas (V1.1.6), "God ... dwells in **heaven** ... ". The closed heaven symbolizes 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:
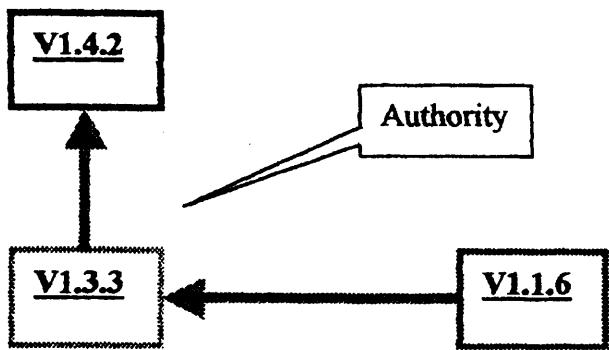
V1.1.6 and V1.3.3 involve a change of Visionary Character but the idea of listening to the words uttered by the character is present. The first character commands Hermas to listen to the words she is going to speak, the second character commands Hermas to listen to the words she is going to read. Hermas 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.
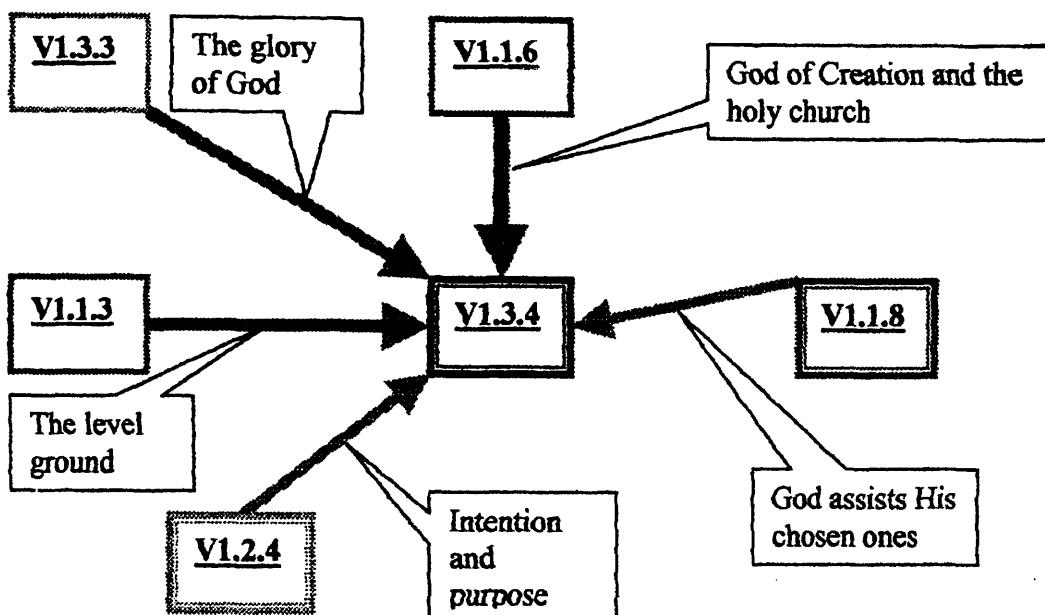
The callout boxes summarise the theme of the reference bonds. V1.1.6 forms the strongest bond with V1.3.4, the latter being a more elaborate statement of the former. The terms 'holy church' and 'creation' are the major terms in this bond. The church is God's creation and is specially blessed. Similarly, 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[120] 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.

---

120   Martin, J.R. "Appraisal Systems in English" in Hunston, S.& Thompson, G. (eds.) *Evaluation in Text.* p.142-175

1. *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, ἱλαρός, χαρά. Outward expressions of happiness will include words such as γελάω. Feelings of unhappiness (negative evaluation) may be expressed with words such as:

λυπέω, λυπηρός, κατήφεια, κατηφής, 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" (V1.1.2). 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** (V1.2.2) 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, (V1.3.4), 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 (V2.4.1).

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[121] 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 1*. 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' (Vl.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 standaradize 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.

Growth of Links with Link Strength

$$y = -0.0191x^6 + 1.406x^5 - 42.542x^4 + 675.83x^3 - 5926x^2 + 27053x - 49775$$
$$R^2 = 0.9977$$
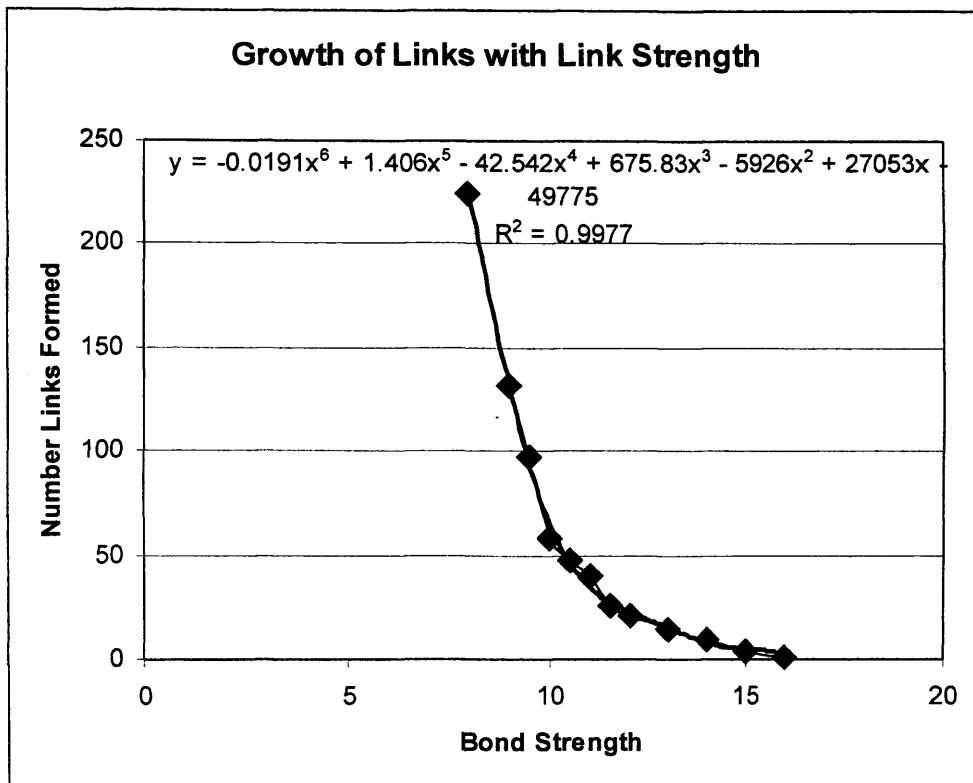
The frequency of the links in the data is shown below:

Link Frequency
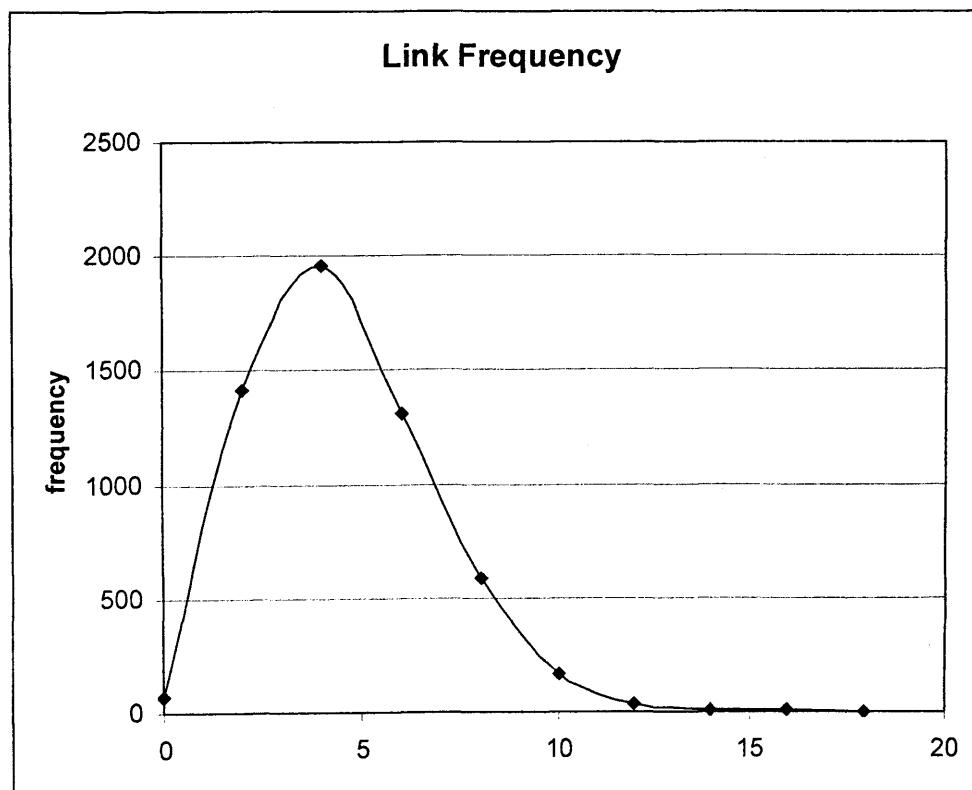
If we look at very strong links first, and take anything ≥ 12.0 as being a very

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.
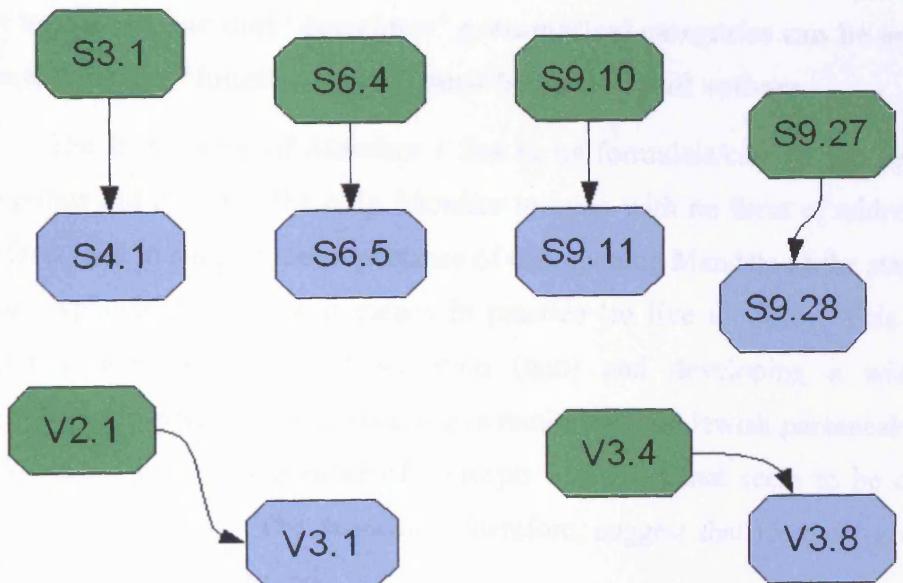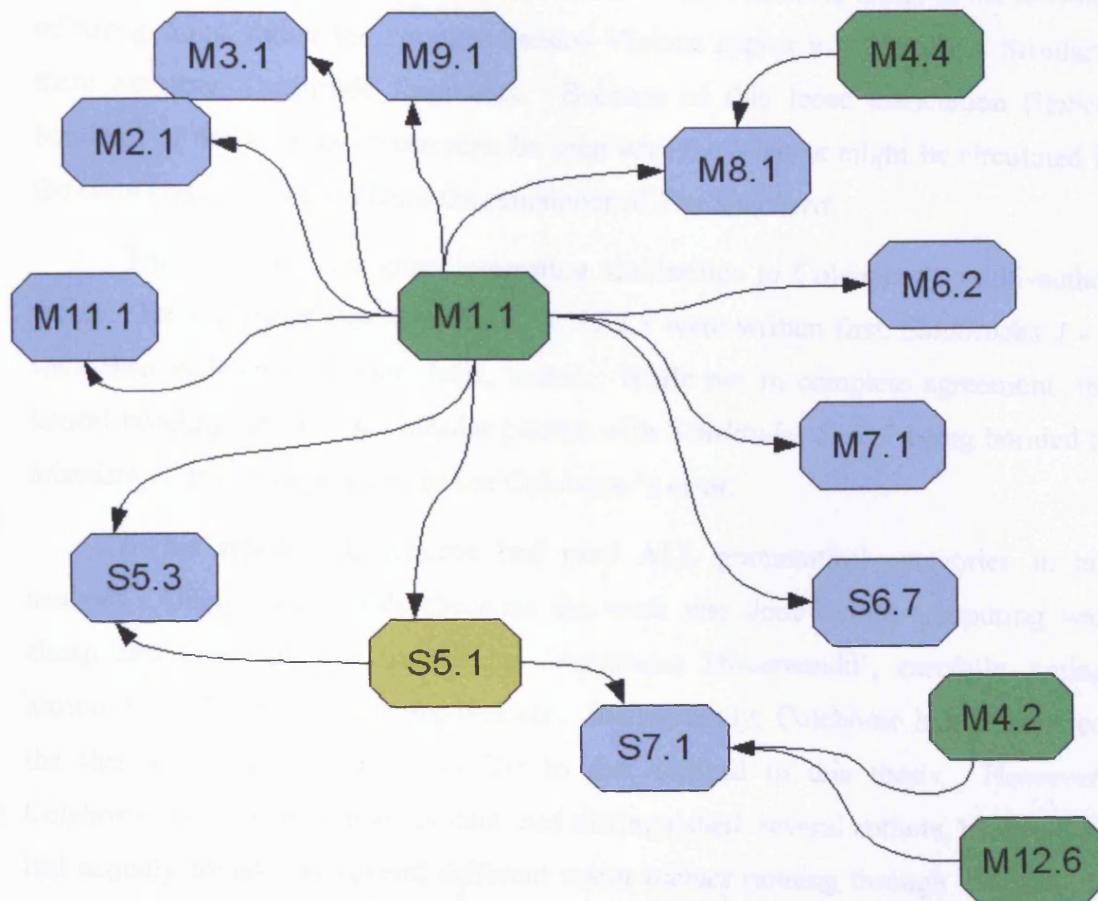
Where Mandates 4 and 12 are found in the network, they are linked only

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 Discernendii', 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[122] 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.

M12.6

V3.1

V2.1

V2.4

M3.1

S7.1

M11.1

M8.1

S5.1

M12.3

S6.1

M4.2

M7.1

V3.4

V3.8

S5.3

S8.6

M4.3

S6.4

S9.14

S5.4

S5.5

S6.5

S9.10

S9.11

S3.1

S4.1

S9.3

S9.4

S9.27

S9.28
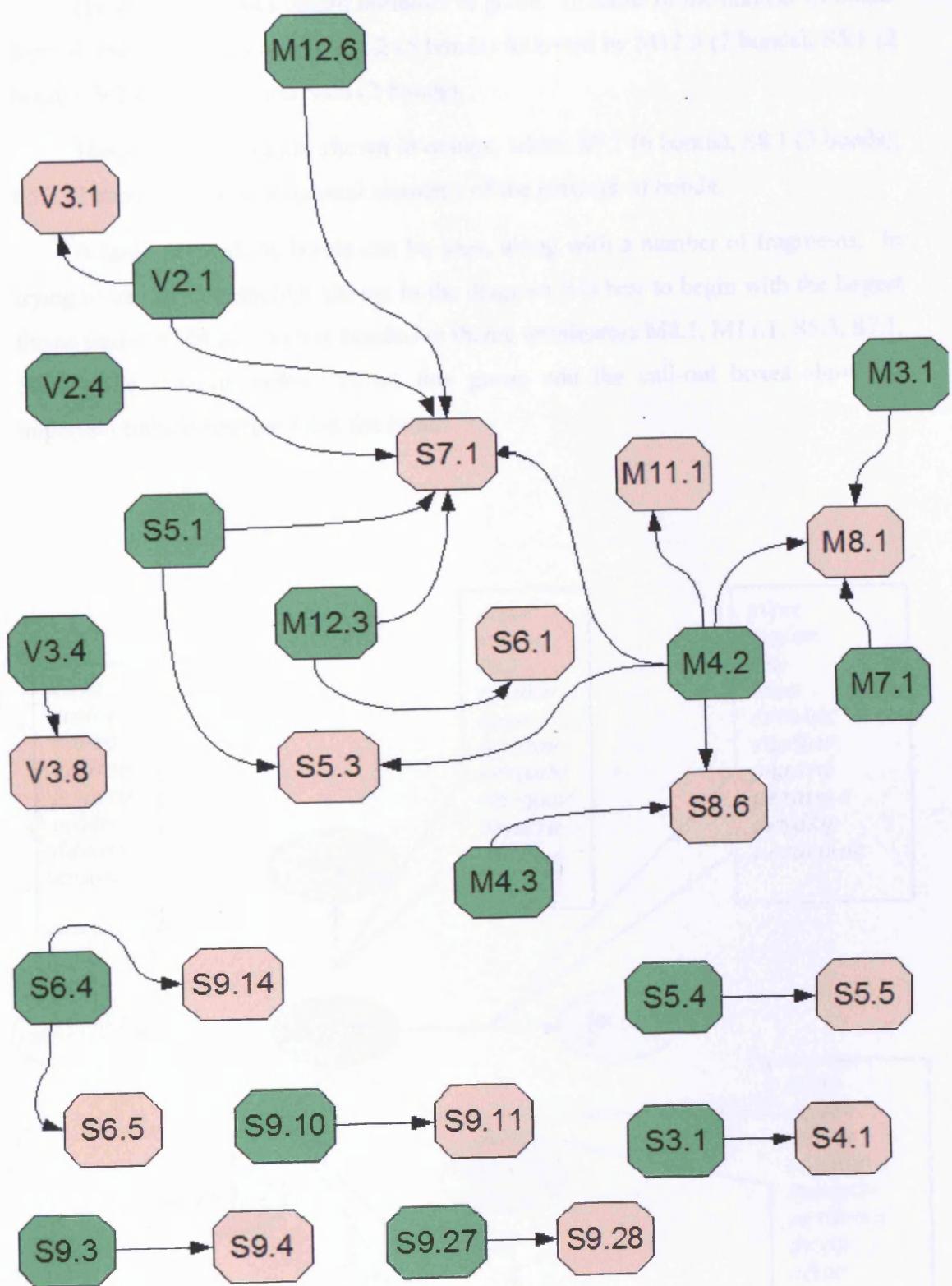
The diagram shows theme initiators in green. In terms of the number of bonds formed, the most important is M4.2 (5 bonds) followed by M12.3 (2 bonds), S5.1 (2 bonds), V2.1 (2 bonds) and S6.4 (2 bonds).

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.
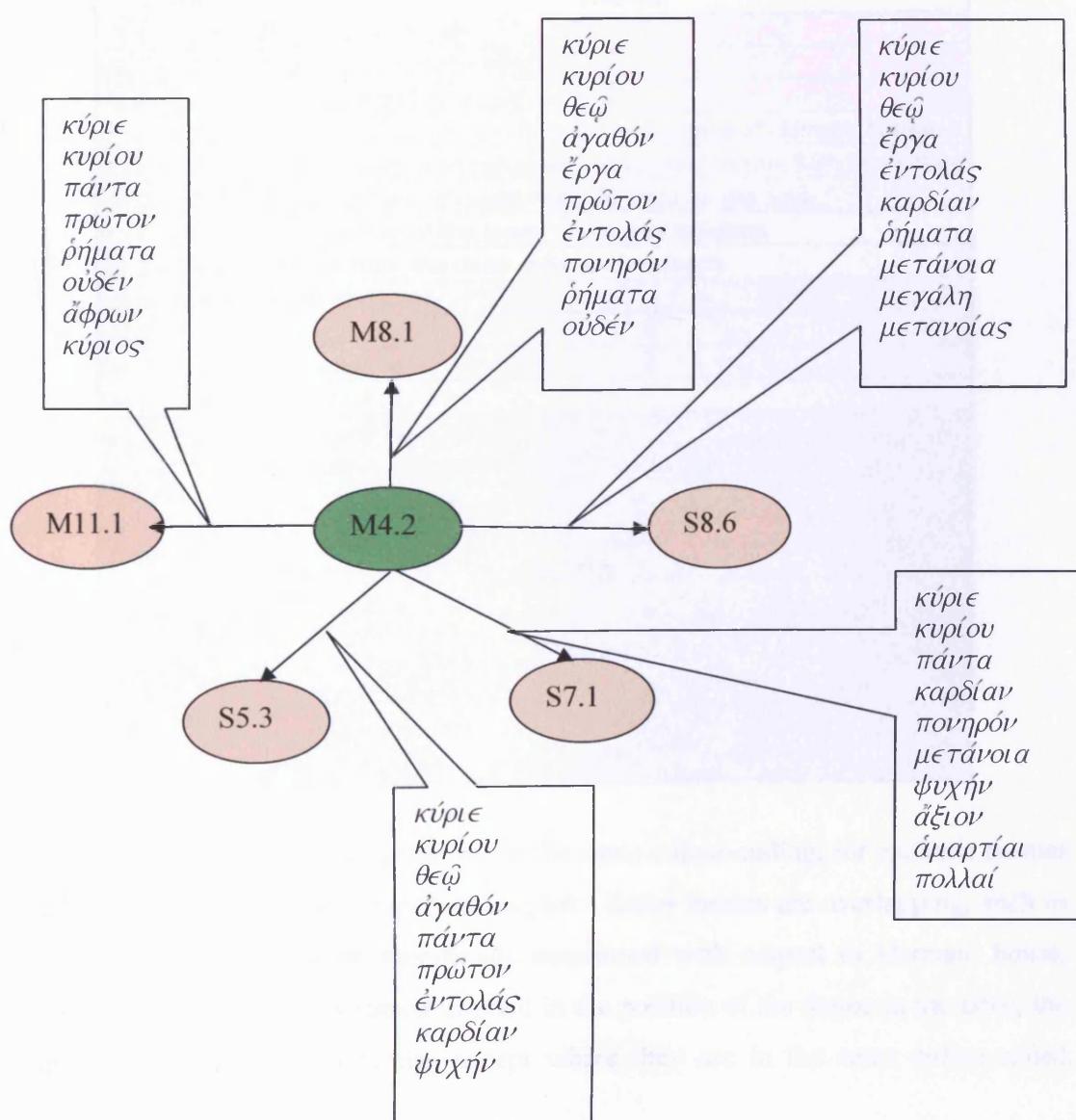
There are some words common to all the bonds *(κύριε* and *κυρίου)* and some words that are characteristic of some bonds only. We can see that M4.2 bonds with M11.1 by the word *ἄφρων*. Therefore there is a link around the theme of 'foolishness'. Proceeding in this manner we can find themes concerning '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:

Bond	Theme
V2.1 - V3.1	Hermas at prayer.
V2.1 - V7.1	Hermas thought worthy.
V2.4 - S7.1	Hermas and his house.
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 is good for the soul.
V3.4 - V3.8	The building of the tower : men and maidens.
S9.3 - S9.4	Stones from the deep, men and maidens.
M3.1 - M8.1	Truth and Lies.
M4.2 - S7.1	Repentance.
M4.3 - S8.6	Repentance.
M4.2 - S8.6	Repentance.
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.
S5.4 - S5.5	Parable of vineyard.
S9.27 - S9.28	Image of trees and fruitfulness.
M4.2 - M11.1	Foolishness.
S6.4 - S9.14	Foolishness.
S6.4 - S6.5	Foolishness of luxury and deceit.
S9.10 - S9.11	Cheerfulness.
S3.1 - S4.1	Withered trees in winter.
S5.1 - S5.3	Acceptable fast.

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

group.

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[123] accepted single authorship, and used the text of *The Shepherd* as a source of biographical information about Hermas. Dibelius[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[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[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* <u>does</u> 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

123   Wilson, W.J. "The career of the Prophet Hermas". *HTR* (1927) 20, p.21-62.
124   Dibelius, M. *Der Hirt des Hermas.*
125   Lampe, P. *From Paul to Valentinus.* pp.227-236.
126   Audet, J.P. "Affinités littéraires et doctrinales du Manuel de Discipline". R.B. (1953) 60, pp.41-82. Quoted in Reiling,J. *Hermas and Christian Prophecy.* p.26.

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:

$$ἀρκετή \ σοι \ ἡ \ ὑπόμνησις \ αὕτη \ καί \ τοῖς \ ἁγίοις$$
$$καί \ ἡ ἀνακαίνωσις \ τῶν \ πνευμάτων \ ὑμῶν \ ἀλλ᾽ \ οὐ$$
$$σοί \ μόνῳ \ ταῦτα \ ἀπεκαλύφθη \ ἀλλ᾽ \ ἵνα \ πᾶσιν$$
$$δηλώσῃς \ αὐτά.$$

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

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[128,] and seems to be but one aspect of a wider Jewish-Christian background to *The Shepherd.* Hermas also borrows material from the Hellenistic world, and it is this mixture that Reiling[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 *Mandates* set out the rule of moral and spiritual ethics that the community should keep.

In Mandate12.3.4 Hermas describes the commandments as being 'great and glorious'. Of the 43 times the adjective 'glorious' is used in *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[130] , relying heavily on the work of Geertz[131] and Wilson[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

---

128   Snyder, G. "The Shepherd of Hermas". in Grant, R.M. (ed) *Apostolic Fathers.*
129   Reiling, J. *Hermas and Christian Prophecy.* p.26.
130   Maier, H.O. *The Social Setting of the Ministry.* p.76.
131   Geertz, C. "Ethos, World-View and the Analysis of Sacred Symbols". *Antioch Review* (1958) 17, pp.421-437.
132   Wilson, B.R. *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[133] 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,

M3.1 - M8.1	Truth and Lies.

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 ἀποστερητής (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:

> ἀλλά καί ὅταν ἀκούσωσι περί θεότητος καί
> ἀληθείας ὁ νοῦς αὐτῶν περί τήν πρᾶξιν αὐτῶν
> καταγίνεται καί οὐδέν 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 ταῦτα φυλάξῃς καί ἀπό τοῦ νῦν πᾶσαν
> ἀλήθειανλαλήσῃς δυνήσῃ σεαυτῷ ζωήν
> περιποιήσασθαι Καί ὅς ἄν ἀκούσῃ τήν ἐντολήν
> ταύτην καί 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[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 *(σύνεσις)*. 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"[135]. Perhaps such ideas were brought to the attention of Paul when he writes:

οἵτινες ἀπεθάνομεν τῇ ἁμαρτίᾳ, πῶς ἔτι ζήσομεν ἐν αὐτῇ; (Rom 6:2)

If we accept Bruce's[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:

ἤκουσα φημί κύριε παρά τινων διδασκάλων ὅτι
ἑτέραμετάνοια οὐκ ἐστίν εἰ μή ἐκείνη ὅτε εἰς
ὕδωρ
κατέβημεν καί ἐλάβομεν ἄφεσιν ἁμαρτιῶν ἡμῶν
τῶν προτέρων.

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,

ἐάν δέ ὑπό χεῖρα ἁμαρτάνῃ καί μετανοήσῃ ἁ
σύμφορον ἐστίν τῷ ἀνθρώπῳ τῷ τοιούτῳ δυσκόλω
γάρ ζήσεται
(M4.3.6)

Repentance includes the idea of purity for Hermas, consequently in S7.1.2 his family are called on to repent and purify themselves. Repentance is not an easy momentary affair, it is prolonged:

τῶν οὖν μετανοούντων εὐθύς δοκεῖς τάς ἁμαρτίας
ἀφίεσθαιοὖ παντελῶς ἀλλά δεῖ τόν μετανοοῦντα
βασανίσαι τήν ἑαυτοῦ ψυχήν καί
ταπεινοφρονῆσαι ἐν πάσῃ πράξει αὐτοῦἰσχυρῶς

---

135   Lampe, P. *From Paul to Valentinus.* p.94.
136   Bruce, F.F. "To the Hebrews: A Document of Roman Christianity?" *ANRW Part II: Principate* vol. 25.4 Walter De Gruyter; Berlin (1987) pp. 3496-3601.

καί θλιβῆναι ἐν πάσαις θίψεσι ποικίλαις.
(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 ταπεινοφρονῶν καί λειτουργῶν
τῷ κυρίῳ ἐν πάσῃ καθαρᾷ καρδίᾳ ... καί πορεύου
ἐν ταῖς ἐντολαῖς ... αἱ δυνήσεται σου ἡ
μετάνοια ἰσχυρά καί καθαρά εἶναι.
(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 ἀπώλοντο 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 θέλῃς τό πονηρόν ἐργάσασθαι φοβοῦ τόν
κύριον καί οὐκ ἐργάσῃ αὐτό·
ἐάν δέ θέλῃς πάλιν τό ἀγαθόν ἐργάσασθαι φοβοῦ
τόν κύριον καί ἐργάσῃ αὐτό ·

The sections from the *Similitudes* link the doing of good with purity of heart, so:

> μηδέν πονηρεύσῃ ἐν τῇ ζωῇ σου ἀλλά δούλευσον
> τῷ κυρίῳ ἐν καθαρᾷ καρδίᾳ.
> (S5.1.5)

S5.4 - S5.5	Parable of vineyard.
S9.27 - S9.28	Image of trees and fruitfulness.

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 γνωστός (known). The important verses of these sections then are S5.4.2&3 and S5.5.1:

> ὁ δέ ἀποκριθείς μοι εἶπεν αὐθάδης εἶ λίαν εἰς τό
> ἐπερωτᾶν ... ὅς ἄν δοῦλος ᾖ τοῦ θεοῦ καί ἔχῃ τόν κύριον
> ἑαυτοῦ ἐν τῇ καρδίᾳ αἰτεῖται παρ᾽ αὐτοῦ σύνεσιν καί
> λαμβάνει καί πᾶσαν παραβολήν ἐπιλύει καί γνωστά αὐτῷ
> γίνονται τά ῥήματα τοῦ κυρίου τά λεγόμενα διά παραβολῶν
> (S5.4.2&3)

> εἶπον σοι φησί καί ἄρτι ὅτι πανοῦργος εἶ καί αὐθάδης
> ἐπερωτῶν τάς ἐπιλύσεις τῶν παραβολῶν ἐπειδή δέ οὕτω
> παράμονος εἶ ἐπιλύσω σοι τήν παραβολήν τοῦ ἀγροῦ
> καί τῶν λοιπῶν τῶν ἀκολούθων πάντων ἵνα γνωστά πᾶσι
> ποιήσῃς αὐτά ἄκουε νῦν φησίκαί σύνιε 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 $(\mathring{\alpha}\phi\rho\omega\nu)$ 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 $(\kappa\epsilon\nu\acute{o}\varsigma)$, 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 $(\tau\rho\upsilon\phi\acute{\eta})$. 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 ἄφρων.

S9.10 - S9.11	Cheerfulness.

These two sections are difficult to understand because of the nature of the material contained in them. Osiek[137] 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' (ἱλαρός). 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 (ὁμαλός), 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 δεδείπνησα ῥήματα κυρίου (S9.11.8).

For Hermas, true joy is found in the service of the church, prayer and reading/teaching.

S3.1 - S4.1	Withered trees in winter.

In this vision Hermas sees trees that are barren of leaves, they all appear to be in the same condition dry/diseased (ξηρά). 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.

S5.1 - S5.3	Acceptable fast.

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 (στατίωνα). This was probably some new practice in the church because the Shepherd is unaware of its meaning[138]. 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.

Καί ταῦτα ἐάν ἐργάσῃ μεγάλην νηστείαν ποιήσεις
καί δεκτήν τῷ θεῷ. (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 Osick, *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.)
140   Schiffrin, D. "Jewish Argument as Sociability", *Language in Society* (1984) 13 pp.311-335

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[141] 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[142] 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[143] 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-

---

141   Rips, L.J. "Reasoning and Conversation". *Psychological Review* (1998) 105 pp.411-441
142   Stein, N.L. & Albro, E.R. "The Origins and Nature of Arguments: Studies in Conflict Understanding, Emotion and Negotiation". *Discourse Processes* (2001) 32 pp.113-133
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 Schiffrin'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. Esler[145] regards models as part of the human process of understanding and he rejects[146] the claim made by Horrel that his view of human culture is that human behaviour is predictable and regular.

Elliott[147] 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

---

144   Horrell, D.G. "Models and Methods In Social-Scientific Interpretation: A Response to Philip Esler". *Journal For The Study Of The New Testament* (2000) 78 p.93
145   Esler, P.F. (ed) "Intoduction: Models, Context and Kerygma in New Testament Interpretation" in *Modelling Early Christianity.*
146   Esler, P.F. "Models In New Testament Interpretation: A Reply to David Horrell." *Journal for the Study of The New Testament.* (2000) 78 pp.107-113
147   Elliott, J. *Social-Scientific Criticism of The New Testament.*

are dependent on Carney[148] who argued:

> *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 **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.

Linguistic Analysis Of Conflict

The linguistic mechanisms used by writers[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 *the linguistic method that a participant in argumentation uses in order to express their relationship to the propositions expressed by the other participants.*[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[151] shows that only a few of all the possible stance markers are regularly used by any individual (out of

---

148   Carney, T.F. *The Shape of The Past: Models and Antiquity.* The reference to this publication is taken from Elliott.
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 *modern* context of conversational argument also.
150   Biber, D. "The Grammatical Marking of Stance, A Cross-Register Comparison". *Journal of Historical Pragmatics* (2004) 5 pp.107-136
151   Precht, K. "Stance Moods in Spoken English: Evidentiality and Affect in British and American Conversation." *Text* (2003) 23 pp.239-257

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[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[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.

d) Indicating the precision of a proposition.

e) Indicating the source of a proposition.

f) Indicating the limitations of a proposition.

Biber[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:

---

152   Biber, D., Johansson, S., Leech, G., Conrad, S., and, Finegan, E. *Longman Grammar of Spoken and Written English.*

153   Conrad, S., Biber, D. "Adverbial Marking of Stance in Speech and Writing" in *Evaluation in Text,* Hunston,S.&ThompsonG. (eds) pp.56-73

154   Biber, D. "The Grammatical Marking of Stance, A Cross-Register Comparison". *Journal of Historical Pragmatics* (2004) 5 pp.107-136

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[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).1 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

---

155 CLAWS5 is an unsupervised grammatical tagging program administered by the Linguistics Department, Lancaster University. www.comp.lancs.ac.uk/ucrel/claws5tags.html

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 higlighted 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

214

'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[156] 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

---

156 Kümmel, W.G., *Introduction To The New Testament.* See note on p.287

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 ἐάν. 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 ἔξεστι).

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[157] 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.

REF	STATEMENT	CONFLICT
01:11	ἐδηλώθη γάρ μοι --- ὅτι ἔριδες ἐν ὑμῖν εἰσιν	Conflict initiation: Divisions
01:14	εὐχαριστῶ (τῷ θεῷ) ὅτι οὐδένα ὑμῶν ἐβάπτισα	Baptismal Groups
01:15	ἵνα μή τις εἴπῃ ὅτι εἰς τό ἐμόν ὄνομα ἐβαπτίσθητε	
01:17	οὐ γάρ ἀπέστειλέν με Χριστός βαπτίζειν ἀλλά εὐαγγελίζεσψαι	
01:21	ὁ θεος διά τῆς μωρίας τοῦ κηρύγματος σῶσαι	Wisdom & Boasting
01:26	βλέπετε γάρ --- ὅτι οὐ πολλοί σοφοί	

---

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.

01:29	ὅπως μή καυχήσηται πᾶσα σάρξ ἐνώπιον τοῦ θεοῦ	
02:02	οὐ γάρ ἔκρινά τι εἰδέναι ἐν ὑμῖν εἰ μή Ἰησοῦν Χριστόν	
03:02	γάλα ὑμᾶς ἐπότισα οὐ βρῶμα οὔπω γάρ ἐδύνασθε ἀλλ᾽ οὐδέ ἔτι νῦν δύνασθε	
03:03	ἔτι γάρ σαρκικοί ἐστε ὅπου γάρ ἐν ὑμῖν ζῆλος καί ἔρις	
03:16	οὐκ οἴδατε ὅτι ναός θεοῦ ἐστε	
03:20	καί πάλιν κύριος γινώσκει τούς διαλογισμούς τῶν σοφῶν ὅτι εἰσίν μάταιοι	
04:15	ἐάν γάρ μυρίους παιδαγωγούς ἔχητε ἐν Χριστῷ ἀλλ᾽ οὐ πολλούς πατέρας ἐν γάρ Χριστῷ Ἰησοῦ διά τοῦ εὐαγγελίου ἐγώ ὑμᾶς ἐγέννησα	
04:19	ἐλεύσομαι δέ ταχέως πρός ὑμᾶς ἐάν ὁ κύριος θελήσῃ καί γνώσομαι οὐ τόν λόγον τῶν πεφυσιωμένων ἀλλά τήν δύναμιν	
05:01	ὅλως ἀκούεται ἐν ὑμῖν πορνεία --- ὥστε γυναῖκά τινα τοῦ πατρός ἔχειν	**Conflict initiation: immorality**
05:03	ἐγώ --- ἤδη κέκρικα ὡς παρών τόν οὕτως τοῦτο κατεργασάμενον	**Paul's Attitude to Immoral Persons In The Church**
05:06	οὐκ οἴδατε ὅτι μικρά ζύμη ὅλον τό φύραμα ζυμοῖ	
05:09	ἔγραψα ὑμῖν ἐν τῇ ἐπιστολῇ μή συναναμίγνυσθαι πόρνοις	
05:11	νῦν δέ ἔγραψα ὑμῖν μή συναναμίγνυσθαι ἐάν τις ἀδελφός ὀνομαζόμενό --- τῷ τοιουτῳ μηδέ συνεσθίειν	
05:12	τί γάρ μοι τούς ἔξω κρίνειν	
06:01	τολμᾷ τις ὑμῶν πρᾶγμα ἔχων πρός τόν ἕτερον κρίνεσθαι ἐπί τῶν ἀδίκων	**Lawsuits**
06:02	ἤ οὐκ οἴδατε ὅτι οἱ ἅγιοι τόν κόσμον κρινοῦσιν	
06:03	οὐκ οἴδατε ὅτι ἀγγέλους κρινοῦμεν	
06:04	βιωτικά μέν οὖν κριτήρια ἐάν ἔχητε τούς ἐξουθενημένους ἐν τῇ ἐκκλησίᾳ τούτους καθίζετε	

06:05 οὕτως οὐκ ἔνι ἐν ὑμῖν οὐδείς σοφός ὅς δυνήσεται διακρῖναι ἀνά μέσον τοῦ ἀδελφοῦ αὐτοῦ	
06:07 ἤδη μέν οὖν ὅλως ἥττημα ὑμῖν ἐστιν ὅτι κρίματα ἔχετε μεθ' ἑαυτῶν	
06:09 ἤ οὐκ οἴδατε ὅτι ἄδικοι θεοῦ βασιλείαν οὐ κληρονομήσουσιν	**Immorality And The Body**
06:15 οὐκ οἴδατε ὅτι τά σώματα ὑμῶν μέλη Χριστοῦ ἐστιν --- ποιήσω πόρνης μέλη μή γένοιτο	
06:16 ἤ οὐκ οἴδατε ὅτι ὁ κολλώμενος τῇ πόρνῃ ἕν σῶμά ἐστιν	
06:18 πᾶν ἁμάρτημα ὅ ἐάν ποιήσῃ ἄνθρωπος ἐκτός τοῦ σώματός ἐστιν	
06:19 ἤ οὐκ οἴδατε ὅτι τό σῶμα ὑμῶν ναός τοῦ ἐν ὑμῖν ἁγίου πνεύματός ἐστιν οὗ ἔχετε ἀπό θεοῦ	
07:07 θέλω δέ πάντας ἀνθρώπους εἶναι ὡς καί ἐμαυτόν	**Marriage**
07:08 λέγω δέ τοῖς ἀγάμοις καί ταῖς χήραις καλόν αὐτοῖς ἐάν μείνωσιν ὡς κἀγώ	
07:09 κρεῖττον γάρ ἐστιν γαμῆσαι ἤ πυροῦσθαι	
07:10 γυναῖκα ἀπό ἀνδρός μή χωρισθῆναι	
07:11 ἐάν δέ καί χωρισθῇ μενέτω ἄγαμος ἤ τῷ ἀνδρί καταλλαγήτω καί ἄνδρα γυναῖκα μή ἀφιέναι	
07:21 δοῦλό ἐκλήθης μή σοι μελέτω ἀλλ' εἰ καί δύνασαι ἐλεύθερος γενέσθαι μᾶλλον χρῆσαι	
07:26 νομίζω --- ὅτι καλόν ἀνθρώπῳ τό οὕτως εἶναι	
07:39 ἐάν δέ κοιμηθῇ ὁ ἀνήρ ἐλευθέρα ἐστίν ᾧ θέλει γαμηθῆναι μόνον ἐν κυρίῳ	
08:01 οἴδαμεν ὅτι πάντες γνῶσιν ἔχομεν	**Idols**
08:02 εἴ τις δοκεῖ ἐγνωκέναι τι οὔπω ἔγνω καθώς δεῖ γνῶναι	
08:04 οἴδαμεν ὅτι οὐδέν εἴδωλον ἐν κόσμῳ καί ὅτι οὐδείς θεός εἰ μή εἷς	
09:02 εἰ ἄλλοις οὐκ εἰμί ἀπόστολος ἀλλά γε ὑμῖν εἰμι	**Apostleship**

218

09:04	μή οὐκ ἔχομεν ἐξουσίαν φαγεῖν καὶ πεῖν	
09:05	μή οὐκ ἔχομεν ἐξουσίαν ἀδελφήν γυναῖκα περιάγειν	
09:06	ἤ μόνος ἐγώ καί βαρναβᾶς οὐκ ἔχομεν ἐξουσίαν μή ἐργάζεσθαι	
09:13	οὐκ οἴδατε ὅτι οἱ τά ἱερά ἐργαζόμενοι τά ἐκ τοῦ ἱεροῦ ἐσθίουσιν	
09:14	τοῖς τό εὐαγγέλιον καταγγέλλουσιν ἐκ τοῦ εὐαγγελίου ζῆν	
09:15	γάρ μοι μᾶλλον ἀποθανεῖν ἤ τό καύχημά μου οὐδείς κενώσει	
09:16	ἐάν γάρ εὐαγγελίζωμαι οὐκ ἔστιν μοι καύχημα ἀνάγκη γάρ μοι ἐπίκειται οὐαί γάρ μοί ἐστιν ἐάν μή εὐαγγελίσωμαι	
09:24	οὐκ οἴδατε ὅτι οἱ ἐν σταδίῳ τρέχοντες πάντες μέν τρέχουσιν εἷς δέ λαμβάνει τό βραβεῖον	
09:26	ἐγώ τοίνυν οὕτως τρέχω ὡς οὐκ ἀδήλως	
10:01	οὐ θέλω γάρ ὑμᾶς ἀγνοεῖν --- ὅτι οἱ πατέρες ἡμῶν πάντες ὑπό τήν νεφέλην ἦσαν καί πάντες διά τῆς θαλάσσης διῆλθον	Idolatry
10:21	οὐ δύνασθε ποτήριον κυρίου πίνειν καί ποτήριον δαιμονίων οὐ δύνασθε τραπέζης κυρίου μετέχειν καί τραπέζης δαιμονίων	
10:28	ἐάν δέ τις ὑμῖν εἴπῃ τοῦτο ἱεροθυτ όν ἐστιν μή ἐσθιετε	
11:22	μή γάρ οἰκίας οὐκ ἔχετε εἰς τό ἐσθίειν καί πίνειν --- τί εἴπω ὑμῖν ἐπαινέσω ὑμᾶς ἐν τούτῳ οὐκ ἐπαινῶ	
12:01	περί δέ τῶν πνευματικῶν ἀδελφοί οὐ θελω ὑμας ἀγνοεῖν	Charismata
14:05	θέλω δέ πάντας ὑμᾶς λαλεῖν γλώσσαις μᾶλλον δέ ἵνα προφητεύητε	
14:06	νῦν δέ ἀδελφοί ἐάν ἔλθω πρός ὑμάς γλώσσαις λαλῶν τί ὑμάς ὠφελήσω	
14:08	καί γάρ ἐάν ἄδηλον σάλπιγξ φωνήν δῷ τίς παρασκευάσεται εἰς	

	πόλεμον	
14:23	ἐάν οὖν συνέλθῃ ἡ ἐκκλησία ὅλη ἐπί τό αὐτό καί πάντε λαλῶσιν γλώσσαις	
14:28	ἐάν δέ μή ᾖ διερμηνευτής σιγάτω	
14:30	ἐάν δέ ἄλλῳ ἀποκαλυφῇ καθημένῳ ὁ πρῶτος σιγάτω	
14:31	δύνασθε γάρ καθ᾽ ἕνα πάντες προφητεύειν	
14:37	ἐπιγινωσκέτω ἅ γράφω ὑμῖν ὅτι κυρίου ἐστίν ἐντολή	
15:03	παρέδωκα γάρ ὑμίν --- ὅτι Χριστός ἀπέθανεν ὑπέρ τῶν ἁμαρτιῶν	**Paul's Preaching**
15:04	καί ὅτι ἐτάφη καί ὅτι ἐγήγερται τ ῇ ἡμερᾷ τῇ τρίτῃ κατά τάς γραφάς	
15:05	καί ὅτι ὤφψη	
15:12	εἰ δέ Χριστός κηρύσσεται ὅτι ἐκ νεκρῶν ἐγήγερται πῶς λέγουσιν ἐν ὑμίν τινες ὅτι ἀνάστασις νεκρῶν οὐκ ἔστιν	
15:25	δεῖ γάρ αὐτόν βασιλεύειν	
15:50	τοῦτο δέ φημι ἀδελφοί ὅτι σάρξ καί αἷμα βασιλείαν θεοῦ κληρονομῆσαι οὐ δύναται οὐδε ἡ φθορά τήν ἀφψαρσίαν κληρονομεῖ	**Death and Resurrection**
15:53	δεῖ γάρ τό φθαρτόν τοῦτο ἐνδύσασθαι ἀφθαρσίαν	
15:58	εἰδότες ὅτι ὁ κόπος ὑμῶν οὐκ ἔστιν κενός ἐν κυρίῳ	
16:07	οὐ θέλω γάρ ὑμᾶς ἄρτι ἐν παρόδῳ ἰδεῖν ἐλπίζω γάρ χρόνον τινά ἐπιμεῖναι πρός ὑμᾶς ἐάν ὁ κύριος ἐπιτρέψῃ	**Paul's Wish to Visit**

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.

220

- 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[158] about their existence to the struggle between patrons in the church[159]. 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[160] 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[161] 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"[162], 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, οὐκ οἴδατε ὅτι... as highlighted in the tabular analysis above. This must surely

158   Fiorenza, E.S., "Rhetorical Situation and Historical Reconstruction in 1 Corinthians." *NTS* (1987) 33 pp.386-403.
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[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[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[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[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.."[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"[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'"[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.

163　quoted in Chow, K.K. *Patronage and Power*. p.132
164　Horrell, D.G. *Social Ethos*. p.109
165　Horrell, D.G. *Social Ethos*. p.107
166　Horrell, D.G. *Social Ethos*. chapter 5
167　Horrell, D.G. *Social Ethos*. p.177
168　Chow, K.K. Patronage and Power. p.117
169　Horrell, D.G. *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.

| S1 | S2 |
| S3 | S4 |
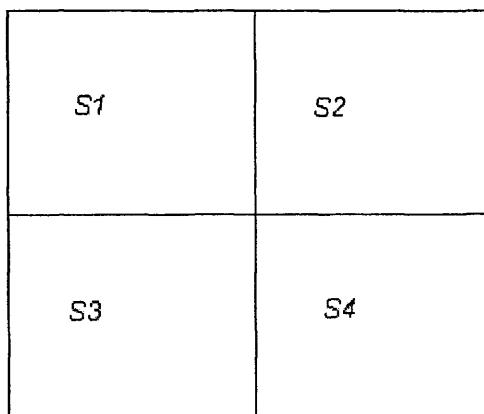
Fig.12.1   *Typical construction of a semiotic square. The symbols are defined in the text.*

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.

THE WORLD Pagan and immoral society	VISIONARY FIGURES Spiritual and other-worldly
HERMAS Mortal with self-doubt	THE CHURCH A society of faith and morality

Fig.12.2   Typical construction of a semiotic square. The symbols are defined in the text.

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 Greimas[170], 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:

ἀλλά πάντως ἐπί τήν καρδίαν σου ἀνέβη περί αὐτῆς (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:

---

170 Greimas,A.J., *Structural Semantics.*

λέγει γάρ σοι δικαίως ὁ κύριος τῆς χώρας ταύτης ἤ τοῖς

νόμοις μου χρῶ ἤ 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[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

---

171 Hunston,S. & Sinclair,J. "A Local Grammar of Evaluation" in *Evaluation in Text*. Hunston,S. &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.

ἄκουε νῦν περί τῶν λίθων τῶν ὑπαγόντων εἰς τήν
οἰκοδομήν. οἱ μέν οὖν λίθοι οἱ τετράγωνοι καί
λευκοί καί συμφωνοῦντες ταῖς ἁρμογαῖς αὐτῶν οὗτοι
εἰσίν οἱ ἀπόστολοι καί ἐπίσκοποι καί διδάσκαλοι καί
διάκονοι οἱ πορευθέντες κατά τήν σεμνότητα τοῦ
θεοῦ καί ἐπισκοπήσαντες καί διδάξαντες καί
διακονήσαντες ἁγνῶς καί σεμνῶς τοῖς ἐκλεκτοῖς τοῦ
θεοῦ οἱ μέν κεκοιμημένοι οἱ δέ ἔτι ὄντες· Καί
πάντοτε ἐαυτοῖς συνεφώνησαν καί ἐν ἑαυτοῖς εἰρήνην
ἔσχον καί ἀλλήλων ἤκουσον· διά τοῦτο ἐν τῇ
οἰκοδομῇ τοῦ πύργου συμφωνοῦσιν αἱ 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:

Καί πάντοτε ἐαυτοῖς συνεφώνησαν
καί ἐν ἐαυτοῖς εἰρήνην ἔσχον
καί ἀλλήλων ἤκουσον·

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:

ἀπόστολοι καί διδάσκαλοι οἱ κηρύξαντες εἰς ὅλον τόν κόσμον καί οἱ διδάξαντες σεμνῶς καί ἀγνῶς τόν λόγον τοῦ κυρίου καί μηδέν ὅλως νοσφισάμενοι εἰς ἐπιθυμίαν πονηράν ἀλλά πάντοτε ἐν δικαιοσύνη καί ἀληθεία πορευθέντες καθώς καί παρέλαβον τό πνεῦμα τό ἅγιον. τῶν τοιούτων οὖν ἡ πάροδος μετά τῶν ἀγγέλων ἐστίν (S9.25.2)

Hermas evaluates the preaching and teaching of the Apostles and teachers as being proclaimed reverently *(σεμνῶς)* and purely *(ἀγνῶς)*. He then remarks of these that they

μηδέν ὅλως νοσφισάμενοι εἰς ἐπιθυμίαν πονηράν.

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 φιλόξενοι 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 εἰσί κακῶς διακονήσαντες καί διαρπάσαντες χηρῶν καί ὀρφανῶν τήν ζωήν καί ἑαυτοῖς περιποιησάμενοι ἐκ τῆς διακονίας ἧς ἔλαβον διακονῆσαι ἐάν οὖν ἐπιμείνωσι τῇ αὐτῇ ἐπιθυμίᾳ ἀπέθανον καί οὐδεμία αὐτοῖς ἐλπίς ζωῆς ἐάν δέ ἐπιστρέψωσι καί ἁγνῶς τελειώσωσι τήν διακονίαν αὐτῶν δυνήσονται ζῆσαι. (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:

τό δέ χρυσοῦν μέρος ὑμεῖς ἐστε οἱ ἐκφυγόντες τόν κόσμον τοῦτον. ὥσπερ γάρ τό χρυσίον δοκιμάζεται διά τοῦ πυρός καί εὔχρηστον γίνεται οὕτως καί ὑμεῖς δοκιμάζεσθε οἱ κατοικοῦντες ἐν αὐτοῖς. οἱ οὖν μείναντες καί πυρωθέντες ὑπ' αὐτῶν καθαρισθήσεσθε. ὥσπερ τό χρυσίον ἀποβάλλει τήν σκωρίαν αὐτοῦ οὕτω καί ὑμεῖς ἀποβαλεῖτε πᾶσαν λύπην καί στενοχωρίαν καί καθαρισθήσεσθε καί χρήσιμοι ἔσεσθε εἰς τήν 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 ὑπέρ τοῦ ὀνόματος τοῦ υἱοῦ 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:

οἱ δέ ἡμιξήρους ἔχοντες καί ἐν αὐταῖς σχισμάς οὗτοι καί δίψυχοι καί κατάλαλοι εἰσί καί μηδέποτε εἰρηνεύοντες εἰς ἑαυτούς ἀλλά διχοστατοῦντες πάντοτε ἀλλά καί 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.

$$\dot{a}\pi\acute{\epsilon}\chi\epsilon\sigma\theta\alpha\iota \quad o\dot{\vec{v}}\nu \quad \delta\epsilon\hat{\iota} \quad \dot{a}\pi\acute{o} \quad \tau\hat{\omega}\nu \quad \dot{\epsilon}\pi\iota\theta\nu\mu\iota\hat{\omega}\nu \quad \tau\hat{\omega}\nu \quad \pi o\nu\eta\rho\hat{\omega}\nu$$
$$\ddot{\iota}\nu\alpha \quad \dot{a}\pi o\sigma\chi\acute{o}\mu\epsilon\nu o\iota \quad \zeta\acute{\eta}\sigma\eta\tau\epsilon \quad \tau\hat{\omega} \quad \theta\epsilon\hat{\omega} \quad (M12.2.2)$$

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[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,

$$\alpha\ddot{v}\tau\eta \quad o\dot{\vec{v}}\nu \quad \dot{\eta} \quad \lambda\acute{v}\pi\eta \quad \delta o\kappa\epsilon\hat{\iota} \quad \sigma\omega\tau\eta\rho\acute{\iota}\alpha\nu \quad \dot{\epsilon}\chi\epsilon\iota\nu\cdot \quad \ddot{o}\tau\iota \quad \tau\acute{o} \quad \pi o\nu\eta$$
$$\rho\acute{o}\nu \quad \pi\rho\acute{a}\xi\alpha\varsigma \quad \mu\epsilon\tau\epsilon\nu\acute{o}\eta\sigma\epsilon\nu.$$

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 *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

---

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[175], 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[176] 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:

> χοῖνιξ σίτου δηναρίου καὶ τρεῖς χοίνικες κριθῶν
> δηναρίου, καὶ τό ἔλαιον 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[177].

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.*
176 Frank, T(ed). *An Economic Survey of Ancient Rome.*
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[178]. 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.179*

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)[180]. By about 250 AD the reversal of fortunes was complete, the mines were exhausted, food was short, production down and inflation up[181].

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[182] 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.*
181 Alföldy, G. (tr. Braund,G.&Pollock,F.) *The Social History of Rome.* pp.94ff.
182 Lampe, P. *From Paul to Valentinus.* p.139

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:

> ὁ πένης δὲ ἐπιχορηγούμενος ὑπὸ τοῦ πλουσίου
> ἐντυγχάνει τῷ θεῷ εὐχαριστῶν αὐτῷ, ὑπὲρ τοῦ
> διδόντος αὐτῷ· κἀκεῖνος ἔτι ἐπισπουδάζει περὶ
> τοῦ πένητος, ἵνα ἀδιάλειπτος γένηται ἐν τῇ ζωῇ
> αὐτοῦ· οἶδε γάρ ὅτι ἡ τοῦ πένητος ἔντευξις
> 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[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.

---

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'. A corpus of

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 . . .

245

Appendix 3.1
Master Set of Function Words

Count1 ἄχρι , ἄχρις

Count2 ἄφνω

Count3 ἄλλως

Count4 ἄν

Count5 ἄνευ

Count6 ἄντικρυς

Count7 ἄνω , ἄνωθεν

Count8 ἄρτι

Count9 ἄτερ

Count10 ἀδεῶς

Count11 ἀδήλως

Count12 ἀδιαλείπτως

Count13 ἀδίκως

Count14 ἀεί

Count15 ἀφόβως

Count16 ἀφρόνως

Count17 ἀγνῶς

Count18 ἀκριβῶς

Count19 ἀκωλύτως

Count20 ἀληθῶς

Count21 ἀλλαχοῦ

Count22 ἀλλαχόθεν

Count23 ἀλλά , ἀλλ᾽

Count24 ἀμέμπτως

Count25 ἀναντιρρήτως

Count26 ἀνά

Count27 ἀνδρείως

Count28 ἀνόμως

Count29 ἀντιπέρα

Count30 ἀντί , ἀνθ᾽ , ἀντ᾽

Count31 ἀπέναντι

Count32 ἁπλῶς

Count33 ἀποτόμως

Count34 ἀπό , ἀπ᾽ , ἀφ᾽

Count35 ἀπροσωπολήμπτως

Count36 ἀσφαλῶς

Count37 ἀσμένως

Count38 ἀσώτως

Count39 αὔριον

Count40 ἅμα

Count51 δήπου

Count52 διά , δι᾽

Count53 διηνεκῶς

Count54 δικαίως

Count55 διό

Count56 διόπερ

Count57 διότι

Count58 δισσῶς

Count59 δίς

Count60 δίχα

Count61 δυνατῶς

Count62 δυσκόλως

Count63 δωρεάν

Count64 ἑβραϊστί

Count65 ἔμπροσθε , ἔμπροσθεν

Count66 ἔναντι

Count67 ἔνδοθεν

Count68 ἔνθα

Count69 ἔνθεν

Count70 ἔπειτα

Count71 ἔσω , ἔσωθεν

Count72 ἔτι

Count73 ἔξω , ἔξωθεν

Count74 εἵνεκεν

Count75 εἶτα , εἶτεν

Count76 εἰ , εἴπερ , εἴπως

Count77 εἰκῇ

Count78 ἐάν , ἐάνπερ

Count79 ἐχές

Count80 ἐφάπαξ

Count81 ἐγγύς

Count82 ἐκ , ἐξ

Count83 ἐκεῖ , ἐκεῖθεν

Count84 ἐκεῖσε

Count85 ἐκουσίως

Count86 ἐκτενῶς

Count87 ἐναλλάξ

Count88 ἐνδόξως

Count89 ἐνίοτε

Count90 ἐνθάδε

Count101 ἐπέκεινα

Count102 ἐπιμελῶς

Count103 ἐπί , ἐπ᾽ , ἐφ᾽

Count104 ἑτοίμως

Count105 ἐξαίφνης

Count106 ἐξάπινα

Count107 ἐξῆς

Count108 εὖ

Count109 εὖγε

Count110 εὐαρέστως

Count111 εὐκαίρως

Count112 εὐπρεπῶς

Count113 εὐθέως

Count114 εὐθύς

Count115 εὐθύμως

Count116 εὐσεβῶς

Count117 εὐτάκτως

Count118 εὐτόνως

Count119 ἕνεκα

Count120 ἕνεκεν

Count121 ἕως

Count122 φανερῶς

Count123 φιλανθρώπως

Count124 φιλοφρόνως

Count125 φρονίμως

Count126 γάρ

Count127 γέ

Count128 γνησίως

Count129 γοῦν

Count130 ἡδέως

Count131 ἡνίκα

Count132 ἤ , ἤτοι

Count133 ἤδη

Count134 ἤ , ἤπερ

Count135 ἱνατί

Count136 ἴσως

Count137 ἰσχυρῶς

Count138 ἵνα , ἵν᾽

Count139 κἄν

Count140 καί

Count41 ἅπαξ

Count42 βεβαίως

Count43 βραδέως

Count44 χαμαί

Count45 χάριν

Count46 χωρίς

Count47 δεινῶς

Count48 δεῦρο

Count49 δέ

Count50 δή

Count151 καθό

Count152 καθότι

Count153 καθώς

Count154 καθώσπερ

Count155 κατά , κατ' , καθ'

Count156 κατενώπιον

Count157 κατέναντι

Count158 κατωτέρω

Count159 κάτω

Count160 κενῶς

Count161 κομψότερον

Count162 κρυφῇ

Count163 κυκλόθεν

Count164 λαμπρῶς

Count165 λάθρᾳ

Count166 λίαν

Count167 μᾶλλον

Count168 μακρόθεν

Count169 μάλιστα

Count170 μεγάλως

Count171 μενοῦν , μενοῦνγε

Count172 μεντοίγε

Count173 μετά , μετ' , μεθ'

Count174 μεταξύ

Count175 μετέπειτα

Count176 μετρίως

Count177 μέχρι , μέχρις

Count178 μέν

Count179 μέντοι

Count180 μηδαμῶς

Count181 μηδέ , μηδ'

Count182 μηδέποτε

Count183 μηδέπω

Count91 ἐνταῦθα

Count92 ἐντεῦθεν

Count93 ἐντός

Count94 ἐνώπιον

Count95 ἐπαύριον

Count96 ἐπάν

Count97 ἐπάνω

Count98 ἐπειδή

Count99 ἐπειδήπερ

Count100 ἐπεί

Count201 ὁμοθυμαδόν

Count202 ὁμοῦ

Count203 ὁπόταν

Count204 ὁπότε

Count205 ὁσάκις

Count206 ὁσίως

Count207 ὁτέ

Count208 ὄπισθε , ὄπισθεν

Count209 ὀπίσω

Count210 ὀργίλως

Count211 ὀρθῶς

Count212 ὀψέ

Count213 οὔπω

Count214 οὔτε

Count215 οὖν

Count216 οὐ , οὐχ , οὐκ

Count217 οὐχί

Count218 οὐδέ

Count219 οὐδέποτε

Count220 οὐδέπω

Count221 οὐκέτι

Count222 οὐκοῦν

Count223 οὕτω , οὕτως

Count224 ὅλως

Count225 ὅμως

Count226 ὄντως

Count227 ὅπου

Count228 ὅπως

Count229 ὅθεν

Count230 ὅταν

Count231 ὅτι

Count232 πανταχοῦ

Count233 παντελῶς

Count141 καίπερ

Count142 καίτοι

Count143 καίτοιγε

Count144 κἀγώ , κἀμέ

Count145 κἀκεῖ , κἀκεῖθεν

Count146 κακῶς

Count147 καλῶς

Count148 καθά

Count149 καθάπερ

Count150 καθεξῆς

Count251 πνευματικῶς

Count252 ποῖ

Count253 πολλάκις

Count254 πολυμερῶς

Count255 πολυτρόπως

Count256 ποσάκις

Count257 ποτέ

Count258 πού

Count259 πόθεν

Count260 πόρρω , πόρρωθεν

Count261 πότερον

Count262 πρίν

Count263 προθύμως

Count264 πρό

Count265 πρός

Count266 πρωΐς , πρωΐσθεν

Count267 πρώτως

Count268 πῶς

Count269 πώποτε

Count270 θαυμαστῶς

Count271 ῥητῶς

Count272 σαφῶς

Count273 σχεδόν

Count274 σεμνῶς

Count275 σφόδρα

Count276 σήμερον

Count277 σπουδαίως

Count278 συντόμως

Count279 σύν

Count280 σωφρόνως

Count281 ταχέως

Count282 τάχα

Count283 τελείως

Count184 μηκέτι

Count185 μή

Count186 μήν

Count187 μήποτε

Count188 μήπω

Count189 μήπως

Count190 μήτε

Count191 μήτι

Count192 μόγις

Count193 μόλις

Count194 ναί

Count195 νή

Count196 νομίμως

Count197 νουνεχῶς

Count198 νῦν , νυνί

Count199 ὁμοίως

Count200 ὁμολογουμένως

Count234 παραχρῆμα

Count235 παραυτίκα

Count236 παρά , παρ'

Count237 παρεκτός

Count238 πάλαι

Count239 πάλιν

Count240 πάντως

Count241 πάνυ

Count242 περισσοτέρως

Count243 περισσῶς

Count244 περί

Count245 πέλας

Count246 πέραν

Count247 πέρυσι , πέρυσιν

Count248 πικρῶς

Count249 πλήν

Count250 πλουσίως

Count284 τέ

Count285 τοιγαροῦν

Count286 τοί

Count287 τοίνυν

Count288 τότε

Count289 τρίς

Count290 ὑπεράγαν

Count291 ὑπεράνω

Count292 ὑπερέκεινα

Count293 ὑπέρ

Count294 ὑποκάτω

Count295 ὑποκάτωθεν

Count296 ὑπό , ὑπ' , ὑφ'

Count297 ὧδε

Count298 ὥς

Count299 ὡσαύτως

Count300 ὡσεί

Count301 ὥσπερεί

Count302 ὥσπερ

Count303 ὥστε

Appendix 6.1

The first six component loadings of a PCA analysis.

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6
Count111	0.274	-0.165	0.019	0.031	0.005	0.029
Count37	-0.066	0.292	0.129	-0.283	-0.019	-0.053
Count147	-0.121	-0.244	-0.189	-0.096	-0.192	-0.007
Count63	-0.018	0.074	0.029	0.291	0.105	0.319
Count97	-0.254	-0.081	-0.027	-0.179	0.046	0.114
Count211	0.068	0.260	0.191	-0.087	-0.119	0.070
Count137	0.112	0.143	-0.017	-0.141	-0.207	-0.065
Count39	-0.245	-0.019	0.167	0.056	-0.001	0.031
Count109	-0.208	0.104	-0.057	0.285	-0.161	0.117
Count238	-0.084	-0.325	-0.025	0.065	-0.019	0.073
Count18	-0.285	0.058	0.036	0.117	-0.098	0.056
Count60	-0.090	0.048	-0.286	-0.206	-0.121	0.139
Count124	-0.081	-0.141	0.293	-0.136	-0.047	0.067
Count59	-0.247	0.062	0.044	-0.021	-0.033	0.122
Count195	-0.052	0.061	-0.166	-0.014	-0.050	-0.380
Count103	-0.188	-0.036	-0.073	-0.251	0.083	0.242
Count178	-0.148	-0.213	-0.098	-0.159	-0.174	-0.064
Count187	-0.027	0.095	-0.264	-0.126	-0.157	0.014
Count236	-0.211	0.004	0.064	-0.193	0.144	-0.080
Count157	-0.123	0.167	0.188	0.029	-0.202	-0.122
Count191	-0.073	0.030	0.093	-0.199	0.021	-0.147
Count4	0.001	0.196	-0.146	-0.155	-0.116	0.204
Count142	-0.216	0.077	0.045	-0.156	0.198	-0.158
Count64	0.159	0.208	0.017	-0.080	-0.188	0.127
Count93	0.116	0.234	0.163	-0.202	-0.085	0.186
Count57	0.042	0.084	0.042	-0.027	-0.085	0.065
Count123	-0.240	0.058	0.085	0.172	-0.174	-0.055
Count173	-0.122	-0.001	-0.127	-0.055	0.211	0.266
Count233	-0.191	-0.197	0.117	0.095	-0.234	-0.022
Count190	-0.131	0.169	-0.180	0.197	-0.126	-0.092
Count184	-0.064	0.139	-0.314	-0.006	0.025	-0.013
Count73	0.086	-0.135	-0.031	0.209	0.172	0.140
Count214	-0.201	0.157	-0.219	-0.053	0.092	-0.006
Count171	-0.150	-0.026	-0.110	-0.083	0.273	0.254
Count223	-0.073	0.057	0.312	-0.039	0.173	-0.168
Count21	0.134	0.021	0.174	-0.131	-0.170	0.174

Count242	-0.139	0.026	0.215	-0.018	0.005	0.103
Count230	-0.044	0.070	-0.028	-0.033	0.163	-0.268
Count210	0.037	-0.077	-0.066	-0.129	-0.130	0.172
Count33	-0.068	-0.239	0.097	-0.056	-0.315	0.038
Count114	-0.132	0.199	-0.093	0.301	-0.153	-0.014
Count58	0.015	0.234	-0.028	0.184	-0.069	-0.009
Count149	0.081	-0.188	-0.167	-0.143	-0.273	-0.188
Count166	0.127	-0.011	0.035	0.043	-0.172	0.252
Count132	-0.230	-0.089	0.194	0.051	-0.131	0.028

Appendix 7.1

Vector of variable weights obtained by OVW

1 ἄχρι , ἄχρις	0.00036		141 μηδέ , μηδ'	0.00188
2 ἄφνω	0.00005		142 μηδέποτε	0.00000
3 ἄλλως	0.00786		143 μή	0.00001
4 ἄν	0.00119		144 μήν	0.00021
5 ἄνευ	0.00488		145 μήποτε	0.00057
6 ἄνω , ἄνωθεν	0.00357		146 μήτε	0.00128
7 ἄρτι	0.00513		147 μήτι	0.01100
8 ἄτερ	0.01568		148 μόλις	0.00104
9 ἀδιαλείπτως	0.01108		149 ναί	0.00003
10 ἀδίκως	0.00001		150 νή	0.00526
11 ἀεί	0.00536		151 νῦν , νυνί	0.00001
12 ἀφόβως	0.00348		152 ὁμοίως	0.00225
13 ἀφρόνως	0.00898		153 ὁμολογουμένως	0.00058
14 ἀγνῶς	0.00274		154 ὁμοθυμαδόν	0.00883
15 ἀκριβῶς	0.01493		155 ὁμοῦ	0.01372
16 ἀληθῶς	0.00780		156 ὁσάκις	0.00559
17 ἀλλαχόθεν	0.00895		157 ὁσίως	0.01554
18 ἀλλά , ἀλλ'	0.00000		158 ὄπισθε , ὄπισθεν	0.00031
19 ἀμέμπτως	0.00788		159 ὀπίσω	0.00664
20 ἀναντιρρήτως	0.01755		160 ὀργίλως	0.00879
21 ἀνά	0.00007		161 ὀρθῶς	0.00841
22 ἀνδρείως	0.00288		162 ὀψέ	0.00004
23 ἀνόμως	0.00009		163 οὔπω	0.00076
24 ἀντί , ἀνθ' , ἀντ'	0.00007		164 οὔτε	0.00058
25 ἀπέναντι	0.00260		165 οὐ , οὐχ , οὐκ	0.00000
26 ἀπλῶς	0.00824		166 οὐχί	0.00139
27 ἀποτόμως	0.00172		167 οὐδέ	0.00230
28 ἀπό , ἀπ' , ἀφ'	0.00000		168 οὐδέποτε	0.00338
29 ἀσφαλῶς	0.00001		169 οὐδέπω	0.00086
30 αὔριον	0.00114		170 οὐκέτι	0.00110
31 ἅμα	0.00806		171 οὔτω , οὔτως	0.00006
32 ἅπαξ	0.01243		172 ὅλως	0.00040
33 χαμαί	0.00243		173 ὅμως	0.00703
34 χάριν	0.00224		174 ὅπου	0.00701
35 χωρίς	0.00239		175 ὅπως	0.00834

36 δεινῶς	0.00803	176 ὅθεν	0.00520
37 δεῦρο	0.00004	177 ὅταν	0.00110
38 δέ	0.00000	178 ὅτι	0.00001
39 δή	0.00710	179 πανταχοῦ	0.00560
40 διά , δι᾽	0.00000	180 παραχρῆμα	0.00757
41 δικαίως	0.00291	181 παρά , παρ᾽	0.00005
42 διό	0.01385	182 παρεκτός	0.00694
43 διόπερ	0.00602	183 πάλαι	0.00071
44 διότι	0.00700	184 πάλιν	0.00005
45 δισσῶς	0.00971	185 πάντως	0.00000
46 δίς	0.01468	186 περισσοτέρως	0.00909
47 δυνατῶς	0.01493	187 περί	0.00001
48 δυσκόλως	0.00031	188 πέραν	0.01204
49 δωρεάν	0.00195	189 πέρυσι , πέρυσιν	0.00227
50 ἑβραιϲστί	0.00003	190 πικρῶς	0.00031
51 ἔμπροσθε , ἔμπροσθεν	0.00813	191 πλήν	0.00686
52 ἔναντι	0.00324	192 πλουσίως	0.00513
53 ἔνδοθεν	0.00722	193 πνευματικῶς	0.01503
54 ἔνθεν	0.01668	194 πολλάκις	0.00064
55 ἔπειτα	0.00850	195 ποσάκις	0.01699
56 ἔσω , ἔσωθεν	0.00004	196 ποτέ	0.00002
57 ἔτι	0.00291	197 πόθεν	0.00271
58 ἔξω , ἔξωθεν	0.00113	198 πόρρω , πόρρωθεν	0.00002
59 εἰ , εἴπερ , εἴπως	0.00011	199 πότερον	0.00029
60 ἐάν , ἐάνπερ	0.00000	200 πρίν	0.00567
61 ἐφάπαξ	0.00383	201 προθύμως	0.00002
62 ἐγγύς	0.01411	202 πρό	0.01106
63 ἐκ , ἐξ	0.00000	203 πρός	0.00005
64 ἐκεῖ , ἐκεῖθεν	0.00003	204 πρωιϲ , πρωιϲθεν	0.00192
65 ἐκεῖσε	0.00236	205 πρώτως	0.01303
66 ἑκουσίως	0.01136	206 πῶς	0.00152
67 ἐνδόξως	0.00665	207 πώποτε	0.01338
68 ἐνθάδε	0.00456	208 θαυμαστῶς	0.00189
69 ἐνταῦθα	0.00088	209 σχεδόν	0.00092
70 ἐντεῦθεν	0.00163	210 σεμνῶς	0.00032
71 ἐντός	0.00179	211 σφόδρα	0.00540
72 ἐνώπιον	0.00072	212 σήμερον	0.00023
73 ἐπαύριον	0.00049	213 σπουδαίως	0.00294
74 ἐπάν	0.00926	214 σύν	0.00157
75 ἐπάνω	0.00000	215 ταχέως	0.00522
76 ἐπειδή	0.01343	216 τάχα	0.00043

77 ἐπεί	0.00756	217 τοιγαροῦν	0.00039
78 ἐπιμελῶς	0.00403	218 τοίνυν	0.00312
79 ἐπί , ἐπ᾿ , ἐφ᾿	0.00000	219 τότε	0.00574
80 ἐτοίμως	0.00463	220 τρίς	0.01136
81 ἐξαίφνης	0.00050	221 ὑπεράνω	0.00053
82 ἐξάπινα	0.00542	222 ὑπέρ	0.00000
83 ἐξῆς	0.00455	223 ὑποκάτω	0.00949
84 εὐπρεπῶς	0.00574	224 ὑπό , ὑπ᾿ , ὑφ᾿	0.00003
85 εὐθέως	0.00073	225 ωϛδε	0.01559
86 εὐθύς	0.00382	226 ὥς	0.00001
87 εὐσεβῶς	0.00035	227 ὡσαύτως	0.00248
88 εὐτόνως	0.01507	228 ὡσεί	0.00879
89 ἕνεκα	0.00039	229 ὥσπερ	0.00452
90 ἕνεκεν	0.00088	230 ὥστε	0.00003
91 ἕως	0.00000		
92 φανερῶς	0.01764		
93 φιλοφρόνως	0.00004		
94 φρονίμως	0.00186		
95 γάρ	0.00003		
96 γέ	0.00033		
97 γοῦν	0.01032		
98 ἡδέως	0.00627		
99 ἡνίκα	0.00015		
100 ἤ , ἤτοι	0.00002		
101 ἤδη	0.01310		
102 ἤ , ἤπερ	0.00005		
103 ἱνατί	0.00609		
104 ἴσως	0.00908		
105 ἰσχυρῶς	0.01098		
106 ἵνα , ἵν᾿	0.00002		
107 κἄν	0.00240		
108 καί	0.00000		
109 καίπερ	0.01436		
110 καίτοι	0.00418		
111 κἀγώ , κἀμέ	0.00631		
112 κἀκεῖ , κἀκεῖθεν	0.00868		
113 κακῶς	0.01327		
114 καλῶς	0.01138		
115 καθά	0.00853		
116 καθάπερ	0.00057		
117 καθεξῆς	0.01555		

118	*καθό*	0.00001
119	*καθότι*	0.00067
120	*καθώς*	0.00000
121	*κατά , κατ' , καθ'*	0.00000
122	*κατέναντι*	0.00769
123	*κάτω*	0.00256
124	*κενῶς*	0.00170
125	*κυκλόθεν*	0.00396
126	*λαμπρῶς*	0.00112
127	*λάθρα̣*	0.00071
128	*λίαν*	0.00426
129	*μᾶλλον*	0.00079
130	*μακρόθεν*	0.00364
131	*μάλιστα*	0.01028
132	*μεγάλως*	0.00002
133	*μενοῦν , μενοῦνγε*	0.00507
134	*μετά , μετ' , μεθ'*	0.00004
135	*μεταξύ*	0.00001
136	*μετρίως*	0.00000
137	*μέχρι , μέχρις*	0.00045
138	*μέν*	0.00001
139	*μέντοι*	0.00009
140	*μηδαμῶς*	0.01319

Appendix 11.1

Part of Horrell's reply to Esler taken from *Journal for the Study of The New Testament*, (2000), 78, pp.83-84. Part of Speech tagging by CLAWS5.

In_PRP a_AT0 recent_AJ0 issue_NN1 of_PRF The_AT0 Journal_NN1 of_PRF Theological_AJ0 Studies_NN2 ,_, Philip_NP0 Esler_NP0 paid_VVD me_PNP the_AT0 honour_NN1 of_PRF writing_VVG a_AT0 lengthy_AJ0 review_NN1 (_(Esler_NP0 1998a_UNC)_) of_PRF my_DPS book_NN1 ,_, The_AT0 Social_AJ0 Ethos_NN1 of_PRF the_AT0 Corinthian_AJ0 Correspondence_NN1 (_(Horrell_NP0 1996_CRD ;_; hereafter_AV0 SE_NP0)_) ._. In_PRP that_DT0 review_NN1 ,_, rather_CJS21 than_CJS22 focus_VVI on_PRP the_AT0 exegetical_AJ0 sections_NN2 of_PRF the_AT0 book_NN1 ,_, Esler_NP0 chose_VVD 'to_NN1 concentrate_VVB on_PRP the_AT0 theoretical_AJ0 issues_NN2 '_POS which_DTQ were_VBD raised_VVN in_AV021 Part_AV022 I_CRD of_PRF the_AT0 study_NN1 ,_, the_AT0 methodological_AJ0 discussion_NN1 (_(Esler_NP0 1998a_UNC :_: 255_CRD)_) ._. Esler_NP0 '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_AJ0 writers_NN2 ,_, reviewing_VVG the_AT0 state_NN1 of_PRF social-scientific_AJ0 interpretation_NN1 of_PRF the_AT0 New_AJ0 Testament_NN1 ,_, have_VHB highlighted_VVN the_AT0 need_NN1 for_PRP discussion_NN1 of_PRF the_AT0 methodological_AJ0 and_CJC theoretical_AJ0 issues_NN2 that_CJT underlie_VVB the_AT0 exegetical_AJ0 practice_NN1 ._. Susan_NP0 Garrett_NP0 ,_, for_AV021 example_AV022 ,_, insists_VVZ ,_, 'It_VVN is_VBZ ..._... increasingly_AV0 urgent_AJ0 that_CJT scholars_NN2 of_PRF Christian_AJ0 origins_NN2 engage_VVB in_PRP sustained_AJ0 reflection_NN1 on_PRP the_AT0 philosophical_AJ0 implications_NN2 of_PRF the_AT0 perspectives_NN2 and_CJC models_NN2 that_CJT they_PNP choose_VVB to_TO0 employ'_VVI (_(Garrett_NP0 1992_CRD :_: 93_CRD)_) ._. Dale_NP0 Martin_NP0 speaks_VVZ of_PRF 'a_ITJ spectrum_NN1 of_PRF opinion_NN1 about_PRP what_DTQ precisely_AV0 a_AT0 social-scientific_AJ0 method_NN1 should_VM0 be'_VVI ,_, with_PRP much_DT0 of_PRF the_AT0 debate_NN1 ,_, in_PRP his_DPS view_NN1 ,_, centring_VVG 'on_AJ0 the_AT0 use_NN1 of_PRF "_" models_NN2 "_" derived_VVN from_PRP the_AT0 social_AJ0 sciences_NN2 '_POS (_(Martin_NP0 1993_CRD :_: 107_CRD)_) ._. Furthermore_AV0 ,_, like_PRP Garrett_NP0 ,_, Martin_NP0 maintains_VVZ that_CJT the_AT0 debate_NN1 reflects_VVZ 'profound_NN1 theoretical_AJ0 differences_NN2 ,_, particularly_AV0 epistemological_AJ0 ones_NN2 '_POS (_(Martin_NP0 1993_CRD :_: 109-10_CRD)_) .1_CRD Discussion_NN1 and_CJC reflection_NN1 are_VBB particularly_AV0 important_AJ0 to_TO0 foster_VVI and_CJC promote_VVI dialogue_NN1 among_PRP those_DT0 who_PNQ seek_VVB to_TO0 make_VVI use_NN1 of_PRF the_AT0 social_AJ0 sciences_NN2 in_PRP New_AJ0 Testament_NN1 interpretation_NN1 ,_, especially_AV0 given_VVN the_AT0 sometimes_AV0 rather_AV0 firm_AJ0 division_NN1 between_PRP two_CRD groups_NN2 who_PNQ have_VHB come_VVN to_TO0 be_VBI referred_VVN to_TO0 as_AV0 'the_VVI social_AJ0 historians_NN2 '_POS and_CJC 'the_VVB social_AJ0 scientists_NN2 '_POS ._. In_PRP the_AT0 following_AJ0 essay_NN1 ,_, I_PNP shall_VM0 attempt_VVI to_TO0 respond_VVI to_PRP Esler_NP0 's_POS arguments_NN2 concerning_PRP the_AT0 use_NN1 of_PRF models_NN2 ,_, the_AT0 validity_NN1 of_PRF Berger_NP0 and_CJC Luckmann'_NP0 s_ZZ0 theory_NN1 of_PRF the_AT0 social_AJ0 construction_NN1 of_PRF reality_NN1 ,_, and_CJC ,_, finally_AV0 and_CJC more_AV0 briefly_AV0 ,_, the_AT0 major_AJ0 argument_NN1 of_PRF SE._NP0 1_CRD ._. The_AT0 Use_NN1 of_PRF Models_NN2 ._.

In_PRP SE_NP0 I_PNP took_VVD issue_NN1 with_PRP a_AT0 widely_AV0 represented_VVN view_NN1 that_CJT a_AT0 social-scientific_AJ0 approach_NN1 to_PRP the_AT0 New_AJ0 Testament_NN1 essentially_AV0 involves_VVZ the_AT0 use_NN1 of_PRF 'models_NN2 ' POS (_(SE_NP0 ,_, 9-18_CRD)_) ._.
I_CRD argued-and_CJC will_VM0 not_XX0 simply_AV0 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_AJ0 method_NN1 and_CJC maintained_VVD that_CJT it_PNP was_VBD certainly_AV0 not_XX0 the_AT0 only_AJ0 way_NN1 to_TO0 approach_VVI the_AT0 New_AJ0 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_AJ0 approach_NN1 can_VM0 lead_VVI to_PRP historically_AV0 and_CJC culturally_AV0 variable_AJ0 evidence_NN1 being_VBG interpreted_VVN through_PRP the_AT0 lens_NN1 of_PRF a_AT0 generalized_AJ0 model_NN1 of_PRF social_AJ0 behaviour_NN1 ._. The_AT0 rich_AJ0 diversity_NN1 of_PRF human_AJ0 behaviour_NN1 is_VBZ thereby_AV0 homogenized_VVN and_CJC explained_VVN in_PRP31 terms_PRP32 of_PRP33 what_DTQ is_VBZ 'typical'-which_DTQ can_VM0 actually_AV0 be_VBI no_AT0 real_AJ0 explanation_NN1 at_AV021 all_AV022 (_(cf._VVB SE_NN1 ,_, 287-89_CRD)_) ._. Although_CJS Esler_NP0 is_VBZ conscious_AJ0 of_PRF these_DT0 dangers_NN2 ,_, and_CJC insists_VVZ on_PRP the_AT0 need_NN1 for_PRP sensitivity_NN1 to_PRP cultural_AJ0 variation_NN1 and_CJC specificity_NN1 (_(see_VVB Esler_NP0 1998b_UNC :_: 4-5_CRD)_) ,_, his_DPS most_AV0 recent_AJ0 work_NN1 still_AV0 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
?	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

CRD	cardinal numeral (e.g. 3, FIFTY-FIVE, 6609) (excl ONE)
DPS	possessive determiner form (e.g. YOUR, THEIR)
DT0	general determiner (e.g. THESE, SOME)
DTQ	wh-determiner (e.g. WHOSE, WHICH)
EX0	existential THERE
ITJ	interjection or other isolate (e.g. OH, YES, MHM)
NN0	noun (neutral for number) (e.g. AIRCRAFT, DATA)
NN1	singular noun (e.g. PENCIL, GOOSE)
NN2	plural noun (e.g. PENCILS, GEESE)
NP0	proper noun (e.g. LONDON, MICHAEL, MARS)
NULL	the null tag (for items not to be tagged)
ORD	ordinal (e.g. SIXTH, 77TH, LAST)
PNI	indefinite pronoun (e.g. NONE, EVERYTHING)
PNP	personal pronoun (e.g. YOU, THEM, OURS)
PNQ	wh-pronoun (e.g. WHO, WHOEVER)
PNX	reflexive pronoun (e.g. ITSELF, OURSELVES)
POS	the possessive (or genitive morpheme) 'S or '
PRF	the preposition OF
PRP	preposition (except for OF) (e.g. FOR, ABOVE, TO)
PUL	punctuation - left bracket (i.e. (or [ )
PUN	punctuation - general mark (i.e. . ! , : ; - ? ...)
PUQ	punctuation - quotation mark (i.e. ` ' ")
PUR	punctuation - right bracket (i.e.) or])
TO0	infinitive marker TO
UNC	"unclassified" items which are not words of the English lexicon
VBB	the "base forms" of the verb "BE" (except the infinitive), i.e. AM, ARE
VBD	past form of the verb "BE", i.e. WAS, WERE
VBG	-ing form of the verb "BE", i.e. BEING
VBI	infinitive of the verb "BE"
VBN	past participle of the verb "BE", i.e. BEEN
VBZ	-s form of the verb "BE", i.e. IS, 'S
VDB	base form of the verb "DO" (except the infinitive), i.e.
VDD	past form of the verb "DO", i.e. DID
VDG	-ing form of the verb "DO", i.e. DOING
VDI	infinitive of the verb "DO"
VDN	past participle of the verb "DO", i.e. DONE
VDZ	-s form of the verb "DO", i.e. DOES
VHB	base form of the verb "HAVE" (except the infinitive), i.e. HAVE
VHD	past tense form of the verb "HAVE", i.e. HAD, 'D
VHG	-ing form of the verb "HAVE", i.e. HAVING
VHI	infinitive of the verb "HAVE"
VHN	past participle of the verb "HAVE", i.e. HAD
VHZ	-s form of the verb "HAVE", i.e. HAS, 'S
VM0	modal auxiliary verb (e.g. CAN, COULD, WILL, 'LL)
VVB	base form of lexical verb (except the infinitive)(e.g. TAKE, LIVE)
VVD	past tense form of lexical verb (e.g. TOOK, LIVED)
VVG	-ing form of lexical verb (e.g. TAKING, LIVING)
VVI	infinitive of lexical verb
VVN	past participle form of lex. verb (e.g. TAKEN, LIVED)
VVZ	-s form of lexical verb (e.g. TAKES, LIVES)
XX0	the negative NOT or N'T
ZZ0	alphabetical symbol (e.g. A, B, c, d)

BIBLIOGRAPHY

Computer Packages

The R Foundation for Statistical Computing, http://www.r-project.org

Packages used:

cluster. version (1.11.0) Martin Maechler, based on S original by Peter Rousseeuw, Anja.Struyf@uia.ua.ac.be and Mia.Hubert@uia.ua.ac.be, and initial R port by Kurt.Hornik@R-project.org.

energy. version (1.0-3) Maria L. Rizzo and Gabor J. Szekely.

mclust. version (2.1-12) C. Fraley and A.E. Raftery, Dept. of Statistics, University of Washington. R port by Ron Wehrens.

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

Abraham, B. & Ledbolter, J. *Statistical Methods For Forecasting.* Wiley & Sons: New York 1983

Alföldy, G. Braund, D. & Pollock, F. (tr.) *The Social History of Rome.* John Hopkins University Press: Baltimore 1988

Attridge, H.W. "Religion Under the Early Empire". *ANRW* II.16.1. De Gruyter: Berlin 1978 pp.45ff.

Aune, D. *Revelation.* (Word Biblical Commentary, 52) Word Books: Dallas 1997

Bahr, G.J. "Paul and Letter Writing in the First Century" *CBQ 28* (1996) pp.465ff.

Barnard, L.W. "The Shepherd of Hermas in Recent Study". *Heythrop Journal 9* (1968) pp.29-36

Barnes, T.D. "Legislation Against the Christians". *JRS 58* (1968) pp.32ff.

Barton, S.C. "The Communal Dimension of Earliest Christanity". *JTS NS 43* (1992) pp.399-427.

Bauckham, R.J. "The Great Tribulation in the Shepherd of Hermas". *JTS 25* (1974) pp.27-40

Bauer, W. Arndt, W.F. & Gingrich, F.W. (tr.) *A Greek-English Lexicon of the New Testament and other Early Christian Literature.* University of Chicago Press: Chicago 1979

Biber, D. Johansson, S. Leech, G. Conrad,S. & Finegan, E. *Longman Grammar of Spoken and Written English.* Longmans: New York 1999

Biber, D. "Historical Patterns for the Grammatical Marking of Stance". *Journal of Historical*

Pragmatics 5/1 (2004) pp.107-136

Biber, D. *Variations Across Speech and Writing.* Cambridge University Press: Cambridge 1988

Binongo, J.N.G.   "Who Wrote the 15th Book of Oz? An Application of Multivariate Analysis to Authorship Attribution". *Chance A Magazine of the American Statistical Association* 16/2 (2033) pp.9ff

Boente, G. Pires, A. & Rodrigues, I.M. "Influence Functions and Outlier Detection Under the Common Principal Components Model: A Robust Approach". *Biometrika* 89/4 (2002) pp.861-875

Bolton, S. *Pharmaceutical Statistics.* Marcel Dekker: New York 1990

Botha, J.E.   "Style in the New Testament: The Need for Serious Reconsideration". *JSNT* 43 (1991) pp.71-87

Breiman, L.   "Bagging Predictors". *Machine Learning* 26/2 (1996) pp.123-140

Brenton, L.C.L.   *The Septuagint with Apocrypha: Greek and English.* Zondervan: Grand Rapids 1992

Briggs, C.L. (ed) *Disorderly Discourse.* Oxford University Press: Oxford 1996

Broughton, T.R.S. *An Economic Survey of Rome.* (4 vols): Baltimore 1938

Brown, R.E. & Meier, J.P. *Antioch and Rome.* London 1982

Bruce, F.F.   "To the Hebrews: A Document of Roman Christianity?" *ANRW* II.25.4 (1987) pp.3496-3601

Burrows, J.F.   "An Ocean Where Each Kind..: Statistical Analysis and Some Major Determinants of Literary Style". *Computers and the Humanities* 23 (1989) pp.309-321

Cameron, A. *Christianity and the Rhetoric of Empire.* University of California Press: Berkeley 1994

Campenhausen von, H. *Ecclesiastical Authority and Spiritual Power in the Church of the First Three Centuries.* Hendrilson: Peabody 1997

Can, F. Patton, J.M. "Change of Writing Style with Time". *Computers and the Humanities* 39 (2004) pp.61-82

Cary, E. (tr) *Dio's Roman History.* Loeb Classical Library Harvard University Press: London 1927

Chaski, C.E.   "Empirical Evaluations of Language-Based Author Identification Techniques". *Forensic Linguistics* 8 (2001)

Clarke, D.D. *Language and Action: A Structuralist Model for Behaviour.* Butterworth: London 1983

Coleborne, W.   "A Linguistic Approach to the Problem of Structure and Composition of the Shepherd of Hermas". *Colloquium* 3 (1969) pp.133-142

Coleborne, W.   "The Shepherd of Hermas: A Case for Multiple Authorship and Some Implications". *Studia Patristica* 10/1 (1970) pp.65-70

259

Collins, J.J *Between Athens and Jerusalem.* Crossroad: New York 1983

Cross, F.L. (ed.) *The Oxford Dictionary of the Christian Church.* Oxford University Press: Oxford 1990

Daley, B.E. *The Hope of the Early Church.* Cambridge University Press: Cambridge 1993

Davison, A.C. & Hinkley, D.V. *Bootstrap Methods and Their Application.* Cambridge University Press: Cambridge 2003

De Soete, G. "OVWTRE: A Program for Optimal Variable Weighting for Ultrametric and Additive Tree Fitting". *Journal of Classification* 5 (1988) pp.101-104

Dibelius, M. *Der Hirt des Hermas.* in Die Apostolischen Väter Mohr Siebeck: Tübingen 1923

Ding, C.S. "Profile Analysis: Multidimensional Scaling Approach". *Practical Assessment, Research & Evaluation* 7 (2001) pp.16ff

Dodds, E.R. *Pagan and Christian in an Age of Anxiety.* Cambridge University Press: Cambridge 1965

Donfried, K.P. & Richardson, P. (eds.) *Judaism and Christianity in First-Century Rome.* Eerdmans: Cambridge 1998

Downing, F.G. "Pliny's Prosecutions of Christians: Revelation and 1 Peter". *JSNT* 34 (1988) pp.105-123

Dugast, D. *Vocabulaire et Stylistique.* Slatkine-Champion: Paris 1979

Dunn, J.D.G. *Unity and Diversity in the New Testament.* SCM Press: London 1994

Elliott, J.H. *A Home for the Homeless: A Sociological Exegesis of 1 Peter.* Fortress Press: Philadelphia 1981

Elliott, J. *Social Scientific Criticism of the New Testament: An Introduction.* SPCK: London 1995

Esler, F.P. "Models in New Testament Interpretation: A Reply to David Horrell". *JSNOT* 78 (2000) 78 pp.107-113

Everitt, B.S. Landau, S. & Leese, M. *Cluster Analysis.* Arnold: London 2001

Fee, G.D. *New Testament Exegesis.* John Knox Press: Ky 1993

Fiorenza, E.S. "Rhetorical Situation and Historical Reconstruction in 1 Corinthians" *NTS* 33 (1987) pp.386-403

Fitzmyer, J.A. *The Gospel According to Luke I-IX.* Anchor Bible Series. Doubleday: Garden City 1981

Fox, R.L. *Pagans and Christians.* Penquin Books: London 1986

Fraley, C. & Raftery, A.E. "MCLUST: Software for Model-Based Clustering, Density Estimation and Discriminant Analysis". Technical Report 415, Department of Statistics: Washington 2003

Frank, T. (ed.) *An Economic Survey of Ancient Rome.* Baltimore 1940

Freeborn, D. *Style, Text Analysis and Linguistic Criticism.* Macmillan: London 1996

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

Garnsey, P.D. "Aspects of the Decline of the Urban Aristocracy". ANRW II.1 (1975) pp.229ff.

Geertz, C. "Ethos, World-View and the Analysis of Sacred Symbols". *Antioch Review* 17 (1958) pp.421-437

Gelderen van, E. *The Rise of Functional Categories.* Benjamins: Amsterdam 1993

Georgakopoou, A. & Groutsos, D. "Conjunctions versus Discourse Markers in Greek: The interaction of Frequency, Position and Functions in Context". *Linguistics* 36/5 (1998) pp.887-917

Gieschen, C.A. *Angelomorphic Christology.* Brill: Leiden

Giet, S. *Hemas et les pasteurs: les trios auteurs du Pasteur d'Hermas* Presses Universitaires de France: Paris 1963

Good, I.J. "The Population Frequencies of Species and the Estimation of Population Parameters". *Biometrika* 40 (1953) pp.237-264

Gordon, A.D. *Classification.* Chapman & Hall: London 1999

Green, H.A. "Power and Knowledge: A Study in the Social Development of Early Christianity". *SR* (1991) pp.20ff

Greimas, A.J. *Structural Semantics. An Attempt at a Method.* University of Nebraska Press: Lincoln 1983

Hahneman, G.M. *The Muratorian Fragment and the Development of the Canon.* Oxford University Press: Oxford 1992

Hair, J.F. Anderson, R.E. Tatham,T.L. & Black, W.C. *Multivariate Data Analysis.* Prentice Hall: Upper Saddle River 1998

Harris, J.R. *Hermas in Arcadia and other Essays.* Cambridge University Press: Cambridge 1896

Hazelton , M.L. "Variable Kernel Density Estimation". *Aust. N.Z. J. Stat.*45/3 (2003) pp.271-284

Henderson, I.H. "Gnomic Quatrains in the Synoptics: An Experiment in Genre Definition". *NTS* 37 (1991) pp.481-498

Henderson, I.H. "Style Switching in the Didache: Fingerprint or Argument?". in Jefford, C.N. (ed) *The Didache in Context: Essays on its Text, History and Transmission.* Brill: Leiden 1995 pp.177-209

Herdan, G. *Quantitative Linguistics.* Butterworth: London 1964

Hoey, M. *Patterns of Lexis in Text.* Oxford University Press: Oxford 1991

Hogben, G.H. "Linguistic Style and Personality". *Language & Style* 10/4 (1977) pp.270ff.

Holton, H. Mackridge, P. & Philippaki-Warburton, I. *Greek: A Comprehensive Grammar of the Modern Language.* Routledge: London 1997

Hopkins, P. "Taxes and Trade in the Roman Empire". *JRS* (1980) pp.101ff.

Horrel, D.G. "Models and Methods in Social-Scientific Interpretation: A response to Philip Esler". *JSNOT* 78 (2000) pp.83-105

Huddleston, R. Hudson, R.A. Winter, E. & Henrici, A. *A Sentence and Clause in Scientific English*. University College London: London 1968

Hunston, S. & Thompson, G. (eds) *Evaluation in Text: Authorial Stance and the Construction of Discourse*. Oxford University Press: Oxford 2001

Johnson, D.E. *Applied Multivariate Methods for Data Analysts*. Duxburry Press: Pacific Grove 1998

Joly, R. *Le Dossier D'Ignace D'Antioche*. Editions De L'Universite De Bruxelles: 1979

Judge, E.A. "The Early Christians as a Scholastic Community". *JRH* 1 (1960) pp.4ff.

Kaufman, L. & Rousseeuw, P.J. *Finding Groups in Data. An Introduction to Cluster Analysis*. Wiley & Sons: New York 1990

Kirkland, A. "The Literary History of The Shepherd of Hermas Visions 1-4". *The Second Sentury: A Journal of Ewarly Christian Studies*. 9/2 (1992) pp.87-102

Kümmel, W.G. *Introduction to the New Testament*. SCM Press: London 1975

Labov, W *Socioloinguistic Patterns*. University of Philadelphia Press: Philadelphia 1972

Lake, K. *The Apostolic Fathers*. (2 vols) Harvard University Press: Cambridge MA 1913

Lake, K. *Eusebius Ecclesiastical History Books I-V*. Harvard University Press: Cambridge MA 2001

Lampe, P. Steinhauser, M. (Tr.) & Johnson, M.D. (Ed.) *From Paul To Valentinus*. Fortress Press: Minneapolis 2003

Lane, W. *Hebrews*. (Word Biblical Commentary,47) Word Books: Dallas 1991

Ledl, T. "Kernel Density Estimation: Theory and Application in Discriminant Analysis". *Austrian Journal of Statistics* 33/3 (2004) pp.267-279

Lightfoot, J.B. *The Apostolic Fathers*. Macmillan: London 1889

Lodhi, H. Saunders, S. Shawe-Taylor,J. Christianini, N. Watkins, C. "Text Classification Using String Kernels". *Journal of Machine Learning Research* 2 (2002) pp.419-444

Longobardi, G. "Reference and Proper Names". *Linguistic Enquiry* 25 (1994) pp.609-666

Lynn, H.S. McCulloch, C.E. "Using Principal Component Analysis and Correspondence Analysis for Estimation in Latent Variable Models". *JASA* 95 (2000) pp.561-574

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

Makarenkov, V. Legendre, P. "Optimal Variable Weighting for Ultrametric and Additive Trees and K-means Partitioning: Methods and Software". *Journal of Classification* 18/2 (2001) pp.245-

Mc Ewan Humphrey, E. *The Ladies and the Cities.* Sheffield Academic Press: Sheffield 1995

McLain, R. "The Problem of 'Style': Another Case in Fuzzy Grammar". *Language & Style* 10/1 (1977) pp.52ff.

Meeks, W.A. *The First Urban Christians.* Yale University Press: New Haven 1983

Mitchell, M.M. & Young, F.M. (eds) *The Cambridge History of Christianity.* Cambridge University Press: Cambridge 2006

Mommsen, T. *The History of Rome: The Provinces From Caesar to Diocletian.* (2 vols): London 1886

Mooney, C.Z. & Duval, R.D. *Bootstrapping: A Nonparametric Approach to Statistical Inference.* a Sage University Paper Sage Publications: London 1993

Morton, A.Q. "The Authorship of Greek Prose". *Journal of the Royal Statistical Society* 128 (1965) pp.169-233

Morton, A.Q. & Michaelson, S. *The Q-sum Plot.* Department of Computer Science, University Edinburgh: Edinburgh 1990

Nanos, M.D. *The Mystery of Romans* Fortress Press: Minneapolis 1996

Neusner, J. (ed) *Christianity, Judaism and other Greco-Roman Cults.:* Leiden 1975

Neyrey, J.H. *An Ideology of Revolt: John's Christology in Social Science Perspective.* Fortress Press: Philadelphia 1988

Noy, D. *Foreigners at Rome.* Duckworth: London 2002

Oakes, P. (ed) *Rome in the Bible and the Early Church.* Paternoster Press: Carlisle 2002

Ong, W.J. *Orality And Literacy: The Technologizing of the Word.* Routledge: New York 1988

Osiek, C. "The Early Second Century Through the Eyes of Hermas". *Biblical Theology Bulletin* 20(1990) pp.116-122

Osiek, C. *The Shepherd of Hermas.* Fortress Press: Minneapolis 1999

Paulsen, H. "Die Bedeultung Des Montanismus Fur Die Herausbildung Des Kanons *VC 32* (1978) pp.19-52

Pernveden, L. *The Concept of Church in The Shepherd of Hermas.* CWK Gleerup: Lund 1966

Polansky, A.M. & Schucany, W. R. "Kernel Smoothing to Improve Bootstrap Confidence Intervals". *J.R. Statist.Soc. (B)* 59/4 (1997) pp.821-838

Porter, S.E. *Idioms of New Testament Greek.* Sheffield Academic Press: Sheffield 1995

Precht, K. "Stance Moods in Spoken English: Evidentiality and Affect in British and American Conversation". *Text* 23 (2003) pp.239-257

Pruscha, H. "Statistical Models for Vocabulary and Text Length With an Application to the New Testament Corpus". *Literary & Linguistic Computing.* 4 (1998) pp.195-198

Radford, A *Syntax: A Minimalist Introduction.* Cambridge University Press: Cambridge 1999

Reiling, J. *Hermas And Christian Prophecy.* E.J.Brill: Leiden 1973 Supplements to Novum Testamentum XXXVII

Richardson, E.C. (tr) *Nicene and post-Nicene Fathers.* (Second Series): 1892

Rips, L.J. "Reasoning and Conversation". *Psychological Review* 105 (1998) pp.411-441

Robbins, V.K. *The Tapestry of Early Christian Discourse: Rhetoric, Society and Ideology.* Routledge: London 1996

Roberts, A. and Donaldson, J. (eds.) , *The Ante-Nicene Christian Library: Translations of the Fathers down to A.D. 325.* T&T Clark: Edinburgh (1866-72)

Robinson, J.M. *Trajectories Through Early Christianity.* 1979

Russell, B. *An Inquiry into Meaning and Truth.* Penquin Books: Harmondsworth 1973

Scenkein, J. *Studies in the Organization of Conversation.* Academic Press: New York 1978

Schiffrin, D. "Jewish Argument as Sociability". *Language in Society* 13 (1984) pp.311-335

Scroggs, R. "The Sociological Interpretation of the New Testament: The Present State of Research *NTS* 26 (1980) pp.164-179

Shaff, P. (ed.) *Nicene and Post-Nicene Fathers.* Christian Classics Library: Grand Rapids (1893) IV "Athanasius".

Sherwin-White, A.N. "The Early Persecutions and the Roman Law Again". *JTS NS* 3 (1952) pp.199-213

Sherwin-White, A.N. *Roman Society and Roman Law in the New Testament* Oxford University Press: Oxford 1963

Shibata, R. "Bootstrap Estimate of Kullback-Leibler Information for Model Selection". *Statistica Sinica* 7 (1997) pp.375-394

Simmel, G. *Conflict: The Web of Group-Affiliations.* Free Press: New York 1955

Stein, N.A. & Albro, E.R. "The Origins and Nature of Arguments: Studies in Conflict, Understanding, Emotion and Negotiation". *Discourse Processes* 32 (2001) pp.113-133

Stephenson, G. *Mathematical Methods for Science Students.* Longmans: London 1968

Stevens, G.L. *New Testament Greek.* University Press of America: Lanham 1995

Stevenson, J. (ed) *A New Eusebius.* SPCK: London 1963

Stewart-Sykes, A. "Hermas the Prophet and Hippolytus the Preacher: The Roman Homily and its Social Context". in Preacher and Audience. Cunningham, M.B. & Allen.P (ed): ()

Stoodley, K.D.C. Lewis, T. & Stainton, C.L.S. *Applied Statistical Techniques.* Wiley & Sons: Chichester 1980

Streeter, B.H. *The Primitive Church.* London 1923

Stuart-Sykes, A. "The Christology of Hermas". *Augustinianum* 37 (1997) pp.273-284

Sundberg, A.C. "Canon Muratori: A Fourth Century List". *HTR* 66 (1973) pp.1ff

Tabachnick, B.G. & Fidell, L.S. *Using Multivariate Statistics.* Allyn & Bacon: Needham Heights 2001

Theissen, G. *The Social Setting of Pauline Christianity: Essays on Corinth.* Fortress Press: Philadelphia 1982

Titscher, S. Meyer, M. Wodak, R. &Vetter, E. Jenner, B.(tr) *Methods of Text and Discourse Analysis.* Sage Publications: London 2000

Trevett, C. *I Have Heard From Some Teachers': The Second-Century Struggle for Forgiveness and Reconciliation.* Boydell Press: 2004

Trevett, C. *Christian Women and the Time of The Apostolic Fathers (AD c.80-160): Corinth, Rome and Asia Minor.* University of Wales Press: Cardiff 2006

Tweedie, F.J. & Baayen, H.R. "How Variable May a Constant be? Measures of Lexical Richness in Perspective". *Computers and the Humanities* 32 (1998) pp.323-352

Van Den Hoek , A. "Techniques of Quotation in Clement of Alexandria: A Review of Ancient Literary Working". *VC* 50 (1996) pp.223-243

Venables, W.N. &Ripley, B.D. *Modern Applied Statistics wth S.* Springer: London 2002

Vlassis, N.A. Papakonstantinou, G. & Tsanakas, P. "Mixture Density Estimation Based on Maximum Likelihood and Sequential Test Statistics *Neural Processing Letters* 9 (1999) pp.63-76

Weisburg, S *Applied Linear Regression.* Wiley & Sons: New York 1980

Whittaker, M. *der 'Hirt des Hermas.* Akademie-Verlag: Berlin 1956

Williams, F. (trs) *The Panarion of Epiphanius of Salamis, Book II and III.* E.J.Brill: Leiden 1993

Wilson, W.J. "The Career of the Prophet Hermas". *HTR* 20 (1927) pp.21-62

Wilson, B.R. *Patterns of Sectarianism. Organization and Ideology in Social and Religious Movements.* Heinemann Educational Books: London 1967

Wilson, J.C. *Five Problems in the Interpretation of the Shepherd of Hermas: Authorship, Genre, Canonicity, Apocalyptic and the Absence of the Name 'Jesus Christ'".* Edwin Mellen Press: Lampeter 1995