1-1-2013

Using Singular Value Decomposition in Classics: Seeking Correlations in Horace, Juvenal and Persius against the Fragments of Lucilius

Thomas Whidden
University of South Carolina

Using Singular Value Decomposition in Classics: Seeking Correlations in Horace, Juvenal and Persius against the fragments of Lucilius

by

Thomas Whidden

Bachelor of Arts
California State University Northridge, 1999

Submitted in Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy in

Comparative Literature

College of Arts and Sciences

University of South Carolina

2013

Accepted by:

Paul Allen Miller, Major Professor

Mark Beck, Committee Member

Catherine Castner, Committee Member

Duncan Buell, Committee Member

Lacy Ford, Vice Provost and Dean of Graduate Studies

DEDICATION

I dedicate this to my beloved wife who smiles and supports me

in my ostensibly eclectic and monetarily unprofitable interests.

ACKNOWLEDGEMENTS

First, I would like to express my sincere appreciation to Dr. Beck who encouraged me many moons ago to actually pursue a degree instead of simply taking random Greek and Latin courses. I wish to thank Dr. Castner whose love for Latin and excitement for the Classics was motivational in my desire to know Latin well. Although Dr. Gardner is not a part of my dissertation committee I have benefited greatly from not only her storehouse of knowledge, but her kindness in letting us graduate students develop our own thoughts about a text even when she knows we are dead wrong. I wish to thank Dr. Sefrin-Weis for her lofty academic standards to push me to excel. The B+ in Aristotle has not only driven me to do better, but has kept me humble. Dr. Buell deserves an award for not only putting up with a hack of a programmer, but he pointed me in the right direction toward SVD. Without this nudge this dissertation would be sorely lacking. I would be remiss unless I credit Dr. Miller who saw this dissertation in its infancy in a shorter paper I wrote for him. His oversight and gracious words of encouragement were greatly appreciated. Last, I wish to credit my wife with her valuable additions and amendations.

ABSTRACT

For the purpose of this dissertation, the hypothesis is posited that a programmatic correlation of the poems of Lucilius and the other Satirists reveals a detailed and dense level of intertextuality, especially in those poems which scholars already understand to allude to the genre's inventor. In addition to those poems which are discussed in secondary literature, we have discovered other poems which correlate highly with the corpus of Lucilius, but have been largely ignored. To demonstrate this fact I have devised a method using Singular Value Decomposition. That method is able to discern this subtle intertextuality in both the texts in question as well as other Classical texts since our method is not language-specific. We have discerned Horace to be the most highly correlated to Lucilius, and further, poem 1.4 to be among the most highly correlated to Lucilius' fragments. In the course of writing this dissertation we will examine other poems which are found to be highly correlated to discover what we hypothesized--if there is a subtle intertextuality which has been largely ignored. We will use what I term a "roving correlation" on target poems to pinpoint dense intertextual areas.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

^{T}	indicates the transposing of a matrix
Σ	indicates summation
ΣUV^T	The factored components of a given matrix using (SVD)

LIST OF ABBREVIATIONS

J.....Juvenal

H.....Horace

P.....Persius

L.....Lucilius

SVD .....Singular Value Decomposition

Introduction

For the purpose of this dissertation, the hypothesis is posited that a programmatic correlation of the poems of Lucilius and the other Satirists reveals a detailed and dense level of intertextuality, especially in those poems that scholars already understand to allude to the genre's inventor. In addition to those poems that are discussed in secondary literature, we have discovered other poems that correlate highly with the corpus of Lucilius, but have been largely ignored. To demonstrate this fact I have devised a method using the known "Singular Value Decomposition algorithm." That method is able to discern this subtle intertextuality in both the texts in question as well as Greek texts since our method is not language-specific. In the course of writing this dissertation we will examine other poems that are found to be highly correlated to discover what we hypothesized, if there is a subtle intertextuality that has been largely ignored. We will use what I term a "roving correlation" (explained below in Chapter 2) on the target poems to pinpoint dense intertextual areas.

In chapter one we introduce digital documents and computer correlations. For the purpose of recognizing the significance of not only the technology of digitizing documents, but the pioneers who were themselves classicists (Roberto Busa and David Packard), a thorough introduction is needed. When one recognizes the profound impact digital documents hold and that every web page one reads is basically a digital document, the world owes the field of classics a

great debt. Since the first two digital works were both Latin, the digitizing of documents has a history based wholly in classical literature.

While these digital documents allowed complete concordances to be easily made, this blazed a trail for further technological advances. One eventual advance is the correlation of documents. A computer correlation is an automatic similarity test using two or more sets of data. We reduce the digital documents in question to sets of data in order to test their similarity. A direct result of pioneers like Busa and Packard who created concordances of Latin works led to further projects in this field discussed below. These projects currently revolutionize how we work today, and also how we learn.

A document correlation of classical texts would be impossible if not for the advances in math in the last two hundred years. A discussion of correlating data would be incomplete without mentioning the inventors of these foundational mathematical methods. Pearson and Galton not only bestowed upon us algebraic gifts, but Pearson makes us acutely aware that correlations can be misleading and therefore we need to be vigilant in interpreting our data. We introduce various methods for comparing documents and then we demonstrate these algorithms in a few simple examples. These simple examples show us the differences and weaknesses of the algorithms introduced and therefore those we should use for our data. We introduce Singular Value Decomposition that looks promising in correlating our documents. We settle upon this algorithm for the basis of our method.

In chapter two we introduce our method. We use the texts of the Roman satirists in a database to do our mathematical correlations. We export the

necessary words ignoring certain common words so that we do not correlate texts based upon insignificant words such as conjunctions, pronouns, etc. (Appendix C). We also create lists of words specific to Roman satire that we use to do specific correlations upon the satirists. We have marked all proper names within the satirists in order to do special proper name correlations. These lists are found in Appendix B. Finally, we prove our method is accurate in identifying similar texts by taking St. Jerome's Latin translation of the Bible to show the Pauline books cluster together. We then use our method to correlate the entire works of each satirist against one another. We use our special satire subject lists against each author as well. We demonstrate that ancient and modern scholarship has shown Horace and *Satire* 1.4 to be the most similar to Lucilius. It is no secret that Horace, Juvenal and Persius all refer to the inventor of their satiric genre, Lucilius. Scholars did not have to make this connection, it was Quintilian, who first comments on the genre of Satire. Quintilian says Lucilius achieved high renown by some, but Horace is "much more polished and pure (10.1.94)." In turn, Horace praises Lucilius as Satire's progenitor. Quintilian says "Satura quidem tota nostra est," "Satire is entirely ours [Roman] (10.1.93; Miller, *Latin Verse Satire* 1)." He cites specifically what Horace says about the style of Lucilius' poems, that they were "a muddy flow out of which you would want to take parts." This is a reference to Horace's *Satire* 1.4.11. Out of the entire genre of satire, it is significant that Quintilian quotes this lone poem to exemplify both the genre and its inventor. Our method confirms Horace to be the highest satirist correlated to Lucilius.

In chapter three we confirm programmatically that *Satire* 1.4 is one of the

highest correlated poems against the books of Lucilius. Since we have confirmed what scholarship has seen with Horace 1.4, we can use these data to find another poem that is highly correlated and do a comparative study on it. This poem should have a dense intertextuality.

In chapter four we perform a comparative study on Juvenal *Satire* 9 and book 26 of Lucilius. We could have used any number of poems from our dataset. The poem in question for the study was selected randomly. We begin with a survey of the scholarship that has been done on Juvenal 9 as well as any scholarship that has compared the *Satires of* Juvenal with those of Lucilius. We determine that this comparative study is unique since scholarship has largely ignored correlating these two texts together. We examine the similarities between both satirists. There is a similar dialectic in each author as well as many didactic aspects. Both also display a negative view of marriage. There are strikingly similar references to Homer. In addition, Lucilius mixes Greek with Latin throughout his *Satires* and Juvenal does this as well in his ninth *Satire*. This exhibits an extremely close likeness to Lucilius. Last, we explore common subjects to Roman satire in each of the documents such as crudeness, sexuality and commerce.

In chapter five we attempt to situate the dubious fragments of Lucilius. There are fragments of Lucilius that are not assigned to any particular book. In as much as we can determine intertextuality accurately, we will also be able to predict to which book the unassigned and dubious fragments of Lucilius belong. We first try to situate unassigned fragments that are known to belong to a subset of Lucilius' books based upon variants in Nonius' text. Next, we offer a

conjecture to situate lines 1196-1208 into book 15 based upon our data and offer an intertextual justification.

In chapter six we conclude by noting the gaps in our data as well as offering suggestions for further research. Finally in chapter seven we describe the tools that have been created in the writing of this dissertation and how they can be used for further research. It is our hope these tools will not only be useful to a few researchers, but may lead to further research.

Chapter 1 - Introduction to Computer Correlation

Given the interdisciplinary nature of this dissertation a thorough introduction to the computer processing of documents is à propos. Without this introduction, the true nature of this dissertation would be impenetrable to the average comparativist or classicist; therefore, some preliminary remarks are necessary. We introduce two areas first because modern document correlation is the product of two independent research ancestries: the history of creating documents in digital form and the history of using mathematical methods to measure similarity. Last, we will demonstrate a few simple examples.

Digital Documents

The process of correlating documents using a computer first starts at digitizing them. A particular document has to be read by a computer in an organized fashion. Instrumental figures like Roberto Busa and David Packard were first to conceive and implement electronic texts in order to create exhaustive concordances. These concordances of Livy and Aquinas are impressive given the rudimentary computer languages of the time as well as the slow nature and memory restrictions on their hardware.

David W. Packard's concordance of Livy that was completed in 1968 was a technologically ground-breaking work, not only because it was one of the first concordances generated by a computer, but because it was the first work to be printed directly to a photo typesetting machine. This work was the fruit of many

long hours in the space of one year by David Packard, who programmed this concordance while at Harvard. These hours were shared by those who spent time typing the text of Livy onto punchcards. One hundred years before David Packard typed the first word of Livy's text onto punchcards, the necessary advances in math and computers had begun. David Packard's work opened the way for other Classical engineering projects.

Oftentimes technology is taken for granted. The Internet is a prime example. If a website lacks a site search (an area that allows a user to search a website for a particular word), the site could be seen as primitive. In like manner, a complete and exhaustive concordance for every work is nowadays a basic necessity. Furthermore, most books that are printed today can be purchased in digital format; this makes them easily searched. With resources like Perseus.org, it is difficult for younger minds to fathom a time when Roman and Greek works lacked an online searchable database, let alone a complete and exhaustive physical concordance. This was the predicament in 1960.

Today there are a variety of programming languages that are powerful, incredibly intuitive and robust in internal functions (Computer). There are many different open source and commercial database systems that make creating indexed works effortless. The most remarkable advances though, that were made, were done so in computer hardware and architecture itself. The speed and storage space today, compared to the sixties, is profound. Handheld phones used today have far more storage and CPU power than could be packed into a computer that took up 1,400 square feet in 1960 (an IBM 701). The cost of one of our phones compared to one of these computers is not even one-tenth of one

percent (compare an iPhone at $500 with an IBM 701 that shipped in 1953 at $1,027,000, Thelen).

It is precisely because the technology in the sixties was so primitive, and because using computers to process classical works was so new, that Packard's concordance was such a monumental feat. At the same time, Packard was being carried on the shoulders of giants with the technology of his day.

Punchcard machines, although seen as primitive today, were a wonder. The punchcard, or the Hollerith card, was named for Herman Hollerith who first conceived the idea to store data on a punchcard that could be read by a machine in 1896 (Punchcard). The original punchcard had been around since 1725 (Punchcard). This card would endure as a reliable storage medium until the early eighties of the 20th century. These cards were used for data storage, and even storage for computer programs. They were stacked in piles of 2000 and read by card readers that would then make their data available to computers.

With the advent of the computer in the early forties, there were men who immediately understood the ramifications of using these machines to manipulate large amounts of data (Computer). Roberto Busa was the first to conceive the idea of creating a concordance with the help of a computer (Winter 4). This Jesuit priest started planning a concordance for the works of Thomas Aquinas in 1946 (Winter 5). This was quite a task as the works of Aquinas exceeded 10,000,000 words. In 1951 he published a work that showed his proof of concept and blazed a trail for others to follow in the ensuing decades (Winter 7). Busa used hand-written punchcards for a single entry of the preposition in (in order to research the clause in his presence) in his proof of concept that would become

one entry of many in his 56 volumes (Winter 6). It would be 20 years before Busa was done typing the works of Aquinas onto punchcards, and 30 years until this voluminous work was finished (Winter 4).

Inspired by Roberto Busa, John Ellison saw the power of what the computer could accomplish. He, Remington Rand, Inc. and a Univac I computer produced a concordance to the Revised Standard Version of the Bible in 1957. This took only a fraction of the time that James Strong took to complete his concordance by hand (Ellison Preface) in the 19th century. Although Roberto Busa is not mentioned specifically in the preface of this Biblical concordance, it is obvious that Busa's contribution to humanities laid the foundation for this work (Winter 4). Ellison used punch tape (almost identical to punchcards, but a continuous strip of paper) that was then transferred to magnetic-tape (Winter 4; Ellison Preface). This is basically the same process that David Packard would use a decade later to generate his own concordance. It is to be noted that technologically at this time, conventional typesetting had to be used. In other words, his computer-generated concordance still needed to be fed into a typesetting machine that would have been a great expense and a hindrance to any humanities departments.

Nothing has been said yet of the advances in computer languages. The concordances mentioned thus far that were conceived by Ellison and Busa, were programmed not by Ellison and Busa themselves, but by professional computer "scientists." I believe one of the reasons why Packard's concordance was completed so quickly was because he was the only programmer on the task; therefore, he didn't have to wait for any sponsoring engineering firm like

Remington Rand, Inc, or IBM (Packard, *A Concordance to Livy: Vol I-IV* vii). It is because of the advances in computer languages that David Packard was able to pick up programming for this project even though his primary training was in Classics (Packard, *A Concordance to Livy: Vol I-IV* vii). Although I am unsure of the language in which Packard programmed (whether FORTRAN, SNOBOL or Assembly), I am sure of two things. First, in 1968, because of the advances making programming more intuitive, David Packard could take on this task of generating a concordance to Livy. And yet, at the same time, comparing the languages available to Packard to the computer languages today, it would have been an insanely tedious process to program an index to any text, let alone Latin, in SNOBOL, FORTRAN, or Assembly language. If Packard had used SNOBOL rather than FORTRAN, his task of creating a concordance would have been less tedious since SNOBOL made it easier to handle strings. All this technology was necessary for David Packard's concordance, both the computer hardware and also the software. The pioneer work by men like Busa and Ellison set the stage for David Packard.

Busa chose Thomas Aquinas because his own dissertation in 1946 was based upon these works. Ellison chose the Bible because he was a man passionate about the Word of God, but why did David Packard choose Livy?

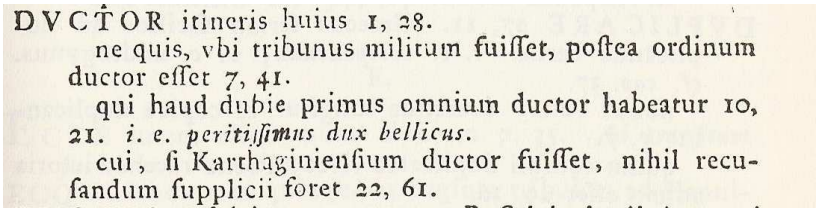
David Packard was studying Classics at Harvard. No doubt, his passion for Classics was profound. This is obvious because the decades after he published his concordance, so much of his time was spent with Greek and Roman works. He would obviously gravitate to a Classical work for this ground-breaking work. A Greek work, although possible in 1966 with punchcards and

software to transfer the text to magnetic-tape (Glickman and Gerrit 1-7), would have been an incredibly arduous task. He would have had to encode a Greek text into the punchcard EBCDIC character set. At this time, the text that could be typed onto a punchcard was limited to what was on a FORTRAN keyboard. This would have been capital A through Z, 0 through 9 and some additional characters (Glickman and Gerrit 24). Further, a Roman work had to be selected that could have been completed within a small time frame since Packard was a fellow for only a year at the Harvard Computing Center. Even though originally Livy was 142 books, in 1968 and to date, we only have 35 books and books 41 and 43 are incomplete (Gould x). One has to wonder if Livy would have been chosen if we had retained all 142 books. It might have taken 3 additional years, or even longer, to type the text onto punch cards. The number of cards would have been multiplied by 4. The number of additional concordance volumes would have easily been 16 with all 142 books. But, as we have only retained 35 books, Livy was small enough to be completed in one year. One last constraint would have been typesetting concerns. It would have been difficult to render the Greek text if one were chosen in place of Livy. One would need to reconfigure the Photon 901 (the typesetting machine Packard used to print his concordance) with a new character set unless he romanized the Greek text. This machine was limited to only 288 characters at a time (Packard, "Publishing Scholarly Compilations by Computer" 75).

Another reason why Livy was chosen was that it serves as a good introduction to all Roman literature. Since Livy, as Gould declares, is a poet through prose, recounting the history of Rome through his own dramatic

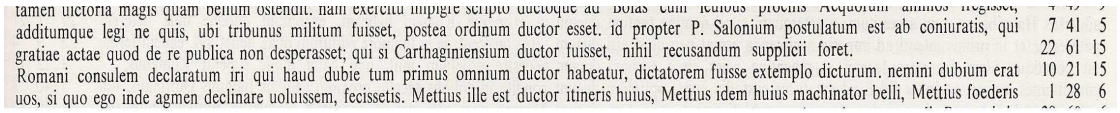
contrivances, what better author to use as the first computer-generated Roman work (Gould xi,xii)? One final reason why Livy was probably chosen was that the only concordance in existence for Livy was the concordance edited by George Olms, and originally published in 1804 by Schäfer and Ernesti. Surely this work took great skill and scholarship to produce without the aid of computers. And this concordance would have been helpful to scholars from 1804 down to the present. But when you compare this concordance to David Packard's concordance, it is sorely obsolete.

Compare this entry from Schäfer and Ernesti's concordance to that of Packard's concordance.



DVCTOR itineris huius 1, 28.
ne quis, vbi tribunus militum fuiſſet, poſtea ordinum ductor eſſet 7, 41.
qui haud dubie primus omnium ductor habeatur 10, 21. *i. e. peritiſſimus dux bellicus.*
cui, ſi Karthaginienſium ductor fuiſſet, nihil recuſandum ſupplicii foret 22, 61.

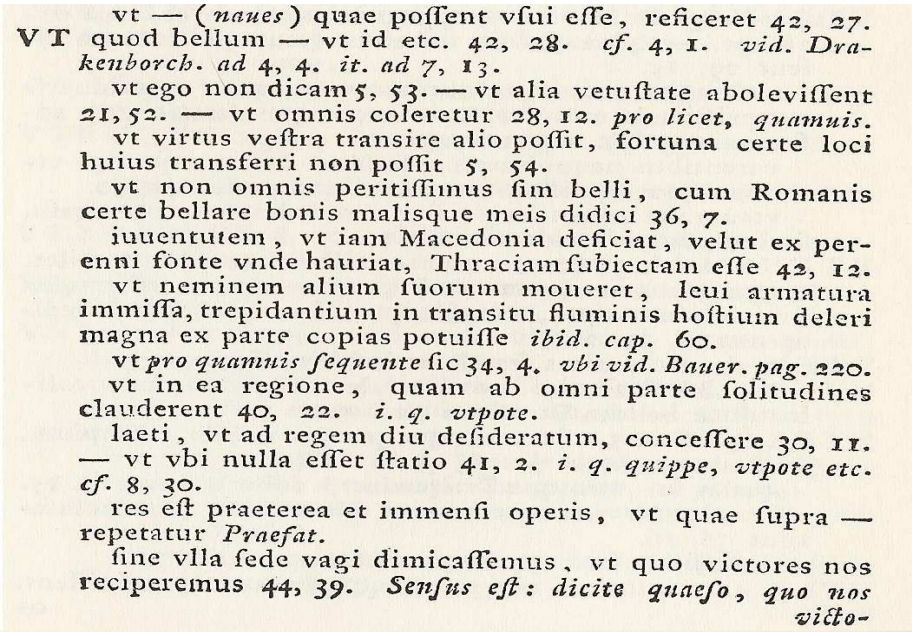
Figure 1.1 Schäfer and Ernesti's concordance.



tamen victoria magis quam bellum ostendit. nam exercitu impigre scripto ductoque ad [illegible] cum [illegible] proelis [illegible] animos [illegible]
additumque legi ne quis, ubi tribunus militum fuisset, postea ordinum ductor esset. id propter P. Salonium postulatum est ab coniuratis, qui 7 41 5
gratiae actae quod de re publica non desperasset; qui si Carthaginiensium ductor fuisset, nihil recusandum supplicii foret. 22 61 15
Romani consulem declaratum iri qui haud dubie tum primus omnium ductor habeatur, dictatorem fuisse extemplo dicturum. nemini dubium erat 10 21 15
uos, si quo ego inde agmen declinare uoluissem, fecissetis. Mettius ille est ductor itineris huius, Mettius idem huius machinator belli, Mettius foederis 1 28 6

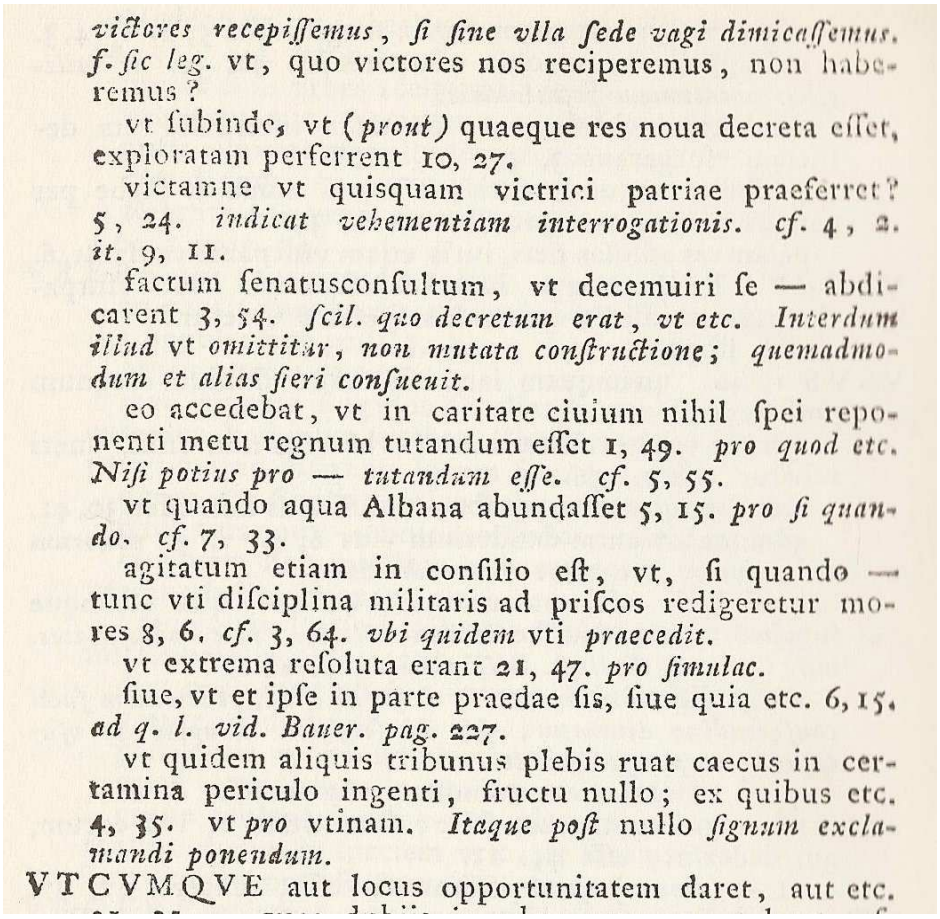
Figure 1.2 Packard's concordance.

While both excerpts have the four entries that exist in Livy, Packard's concordance is easier to read and contains more context before and after the source entry. Furthermore, when we examine a given word with many more occurrences like the common Latin word ut, there is almost no comparison. Here is the complete entry for the Latin word ut from Schäfer and Ernesti.



vt — (*naues*) quae poſſent vſui eſſe, reficeret 42, 27.
VT quod bellum — vt id etc. 42, 28. *cf.* 4, 1. *vid. Drakenborch. ad* 4, 4. *it. ad* 7, 13.
vt ego non dicam 5, 53. — vt alia vetuſtate aboleviſſent 21, 52. — vt omnis coleretur 28, 12. *pro licet, quamuis.*
vt virtus veſtra transire alio poſſit, fortuna certe loci huius transferri non poſſit 5, 54.
vt non omnis peritiſſimus ſim belli, cum Romanis certe bellare bonis malisque meis didici 36, 7.
iuuentutem, vt iam Macedonia deficiat, velut ex perenni fonte vnde hauriat, Thraciam ſubiectam eſſe 42, 12.
vt neminem alium ſuorum moueret, leui armatura immiſſa, trepidantium in transitu fluminis hoſtium deleri magna ex parte copias potuiſſe *ibid. cap.* 60.
vt *pro quamuis ſequente* ſic 34, 4. *vbi vid. Bauer. pag.* 220.
vt in ea regione, quam ab omni parte ſolitudines clauderent 40, 22. *i. q. vtpote.*
laeti, vt ad regem diu deſideratum, conceſſere 30, 11. — vt vbi nulla eſſet ſtatio 41, 2. *i. q. quippe, vtpote etc. cf.* 8, 30.
res eſt praeterea et immenſi operis, vt quae ſupra — repetatur *Praefat.*
ſine vlla ſede vagi dimicaſſemus, vt quo victores nos reciperemus 44, 39. *Senſus eſt: dicite quaeſo, quo nos victo-*

Figure 1.3 Entry 1 for ut from Schäfer and Ernesti's concordance.



victores recepiſſemus, ſi ſine vlla ſede vagi dimicaſſemus. f. ſic leg. vt, quo victores nos reciperemus, non haberemus?
vt ſubinde, vt (*prout*) quaeque res noua decreta eſſet, exploratam perferrent 10, 27.
victamne vt quisquam victrici patriae praeferret? 5, 24. *indicat vehementiam interrogationis. cf.* 4, 2. *it.* 9, 11.
factum ſenatusconſultum, vt decemuiri ſe — abdicarent 3, 54. *ſcil. quo decretum erat, vt etc. Interdum illud* vt *omittitur, non mutata conſtructione; quemadmodum et alias fieri conſueuit.*
eo accedebat, vt in caritate ciuium nihil ſpei reponenti metu regnum tutandum eſſet 1, 49. *pro quod etc. Niſi potius pro — tutandum eſſe. cf.* 5, 55.
vt quando aqua Albana abundaſſet 5, 15. *pro ſi quando. cf.* 7, 33.
agitatum etiam in conſilio eſt, vt, ſi quando — tunc vti diſciplina militaris ad priſcos redigeretur mores 8, 6. *cf.* 3, 64. *vbi quidem* vti *praecedit.*
vt extrema reſoluta erant 21, 47. *pro ſimulac.*
ſiue, vt et ipſe in parte praedae ſis, ſiue quia etc. 6, 15. *ad q. l. vid. Bauer. pag.* 227.
vt quidem aliquis tribunus plebis ruat caecus in certamina periculo ingenti, fructu nullo; ex quibus etc. 4, 35. vt *pro* vtinam. *Itaque poſt* nullo *ſignum exclamandi ponendum.*
VTCVMQVE aut locus opportunitatem daret, aut etc.

Figure 1.4 Entry 2 for ut from Schäfer and Ernesti's concordance.

Schäfer and Ernest chose no more than 35 entries to display from the entire corpus of Livy. This is understandable since the work is only a small single volume. Compare this to Packard's concordance that has every entry for ut. It comprises 47 pages. Packard arranges his concordance entries for every word with subsequent words in alphabetical order so that similar constructions can be easily viewed. This would be helpful for any scholar looking at Livy's use of similar ut constructions. Additionally, the font is so antiquated in Schäfer and Ernest that it is almost unreadable. The references in this old concordance are also difficult to view because they are not lined up. Packard's concordance lines up all entries so that they can be read easily.

Unlike the concordances of Ellison and Busa, David Packard gave us a summary of the process of his concordance, not only in the concordance itself, but in subsequent journal articles. I believe it was this forethought that inspired Humanities departments all over the world to travel along the trail Busa, Ellison and now Packard blazed.

While Packard's work seems like trailblazing through terra incognita, the University of Toronto put out a manual to create concordances of literary works by computers in 1966. In this manual, they outline programs already written in FORTRAN that were specifically for scholars who have no previous knowledge of programming. There are 3 PRORA (Programs for Research On Roman Authors). PRORA I is to transfer a literary text that has been typed out in a certain format on punchcards to magnetic tape, PRORA II prints the text in a certain format, and PRORA III actually creates a rudimentary concordance, or index of the text (Glickman and Gerrit 1-20). While this work, and the

concordances that were created before David Packard's concordance are not mentioned in other computer concordance histories, we have to wonder whether David Packard knew about the research at the University of Toronto. It sounds as though this manual and its corresponding programs (if the University of Toronto would have shared these programs) could have made the work that Packard did infinitely easier.

At any rate, the process that Packard used was not unlike Ellison and Busa. Packard and some others who were attached to Harvard took turns typing out the Oxford classical text of Livy for the first 25 books. After that, the Teubner edition for the remaining 10 books was used. The entire work of Livy took 65,000 punchcards, as it is approximately 505,000 words. This is about 7.8 words per punchcard. When you remember the punchcard machine limitations in characters it will become obvious that conventional typing is not the same as what they had to produce. They would have needed to encode the text somewhat. Brutus alio ratus spectare Pythicam vocem (Gould 88) would need to become, *BRUTUS ALIO RATUS SPECTARE *PYTHICAM VOCEM. Asterisks would need to denote capital letters since all text would be in capital letters. This would add a layer of complexity to the proofing of the text. Additionally, some convention to denote book, chapter and verse would have been needed as well that would create even more noise when proofing the text (Glickman and Gerrit 29).

After the text was completely typed out, these 65,000 punchcards that would have been treated as Vatican-like relics were each proof-read. One person sat at the punchcard machine, that would be whirring much louder than

any computer today, and another would read through the Latin text of Livy. When an error in the punchcard was found, the reading would stop and the offending punchcard or punchcards would be fixed. The old punchcards would then be discarded. This was done through the entire 35 books of Livy (Packard, *A Concordance to Livy: Vol I-IV* vi).

After this first proofing, the 65,000 cards were fed into a card reader attached to an IBM 7094 with magetic tape storage. The cards were read 2000 at a time and a program ran that placed the text of Livy onto magnetic tape. An additional program built an index at this time using the text of Livy and listing every unique word with its references in the entire work of Livy. This is very similar to what the University of Toronto's manual outlines (Glickman and Gerrit 14). It is to be noted that there were no database servers in 1968. Any processing of the text by Packard had to be done manually through a program and stored in some sort of flat text file. This would have been the bulk of the heavy lifting of the concordance. It would have been a feat to do this in any computer language that was around in 1968. Even using a string-friendly language like SNOBOL, Packard would have had to write much logic to create the index in preparation for his concordance. Many languages today are even more conducive to processing text than what would have been available in 1968 (Packard, *A Concordance to Livy: Vol I-IV* vi).

After this index is built another proofing would have been done. They read through each unique word to eliminate erroneous words generated by the index program. It is at this time as well, another program was written to cycle through the text and find any missing chapters or verses in the text that was now on

magnetic tape. Another program as well did a Latin spell check for errors not found during the other proofing stages. Again all this code had to be done from scratch by David Packard. He indicates in his preface that the programming of this concordance took many hours, and often it was done on an irregular schedule to the chagrin of his wife, whom he thanks for her devotion during this time (Packard, *A Concordance to Livy: Vol I-IV* vi).

One last reading aloud was done through the entire text of Livy to find any residual errors. After this last proofing, the program ran that built the actual concordance from the index. Packard mentions that the runtime to build the actual concordance was 3 hours. The concordance to Livy was completed albeit electronically. It existed although only on magnetic tape. They could have printed directly onto an IBM peripheral printer and published it in that form. The typeset would have been abominable and at some point when technology caught up, the concordance would surely have been reprinted. Instead of using a default printer, and because of the costs that would have been involved in conventional typesetting, Packard used a typesetting machine that was able to read the magnetic tape of his concordance attached to the IBM 7094. He mentions that he had to write an additional program in order to print his concordance on the Photon 901 typesetting machine. Essentially, he had to program his own printer driver to finish his concordance. The output from the program that processed his concordance occupied 50 reel to reel tapes, or eleven miles of tape, that is, 133 megabytes (Packard *A Concordance to Livy: Vol I-IV* vi; Packard, "Publishing Scholarly Compilations by Computer" 75). A one gigabyte jump drive today, that is no bigger than your thumb, has 8 times that

storage. Last, Packard mentions that he also created other study tools with the work on magnetic tape, but it was beyond the scope of the concordance (Packard, *A Concordance to Livy: Vol I-IV* vii). It seems at the end of his project, he saw the benefit and potential for even more processing of the text.

Packard's work endures to this day as a useful reference work, and the definitive concordance for the body of Livy's text. It has not been supplanted by another work because it is extremely functional. Aside from the actual concordance, though, because of Packard's work, other departments benefited from his ingenuity in lowering typesetting costs. He inspired other humanities departments to use computer technology in the processing of text. And it also paved the way for other technological projects in the ensuing decades.

First, because Packard's concordance was not supplanted by any other concordance, every work that concerned Livy after 1968 has probably benefited directly from this concordance. It would be unthinkable to write anything on Livy without consulting this exhaustive concordance. Greenaugh in his *Commentary on Livy Books I and II*, that was published in 1976, as well as Gould and Whiteley in their updated edition of *Livy Book I*, published in 1987, no doubt used Packard's concordance to check their own references in their respective prefaces (Gould, xiv,xv; Greenaugh xiv-xvii). In their commentaries, they could easily cross-reference similar clauses to give greater insight to the users of their editions. In addition to these works, any scholarship in Livy would benefit from the use of this concordance.

With the use of the Photon 901, Packard opened the door to other Universities that were cutting costs of not only computer-generated works, but

even for works that are not computer-generated. Any large work that would need to be printed could be typed and sent to a computerized photo typesetting machine. It would obviously have been a huge expense to purchase one of these typesetting machines, but as it could be used for any department, it would swiftly pay for itself. We know Packard's work helped lower costs by his use of typesetting at Harvard and with the Loeb Classical Library years later (Crane).

Finally, Packard's work inspired not only himself for a lifetime of the digital processing of classics, but also motivated other humanities departments to get involved as well. Directly after Packard finished his concordance, he started working on the groundwork of digitizing texts for his Ibycus project. With the Ibycus environment, David Packard modified the Hewlett Packard minicomputer for the optimization of searching digital works (Crane). This environment was purchased by many classics departments all over the world. Although this environment was tailored for classicists, Crane says this project not only inspired the creation of the *Thesaurus Linguae Graecae*, but also influenced their choice of architecture and environment (Crane). David Packard's work also gave Oakman more practical knowledge of computer generated concordances with which to make his recommendations in his manual for building concordances with computers at the University of South Carolina (Oakman 412,413). His work also inspired Howard-Hill to sift through all the various types of approaching computer concordances, and educate any prospective researchers in the area of digital concordances (Howard-Hill 1-4). Scores of other projects had as their inspiration Busa and Packard in the ensuing decades including *Lexicon of Greek Personal Names* (established in 1972 to catalog all Greek names in literature as

a project of Oxford University); *The Gutenberg Project* (purports to have created eBooks in 1971 and seeks to further digitize all books in the public domain); *The Perseus Project* (established in 1985 to allow the reading of Greek and Latin texts online by Tufts University); *Thesaurus Linguae Graecae* (established in 1972 to digitize for search all Greek literature from antiquity to the present as a project of the University of California, Irvine); *Suda On Line: Byzantine* Lexicography (established in 1998 to produce XML-encoded database files of texts); The *Digital Medievalist* (established in 2003 to digitize medieval texts as a project of the University of Lethbridge); *The Homer Multitext Project* (established in 2006 as a project of Harvard University to use digital media to show textual variants not simply in a critical apparatus, but more as an alternate performance of the same story in Homer); *Sermones.net* (established in 2007 to digitize medieval Latin sermons); *Google* Books (established in its infancy in 2002, it partners with libraries and book producers in order to create the largest searchable online library); et al. (Bodard and O'Donnell).

Mathematical Methods to Compare Similarity

The history of using mathematical principles to compare data/documents is more than a hundred years old. These principles are used today as "scientists use bayesian filters to decide if 'this model is better than the alternatives (Hobson, Jaffe, Liddle, Mukherjee and Parkinson 3).'" In *Bayesian Methods in Cosmology*, correlations are used in order to identify extremely remote objects in space. Mathematical methods are used to compare similarity when a Google search is performed, or when Google's news articles are viewed. These news articles have already been run through mathematical filters to predict similarity in

order to group them together. What is astounding is that the formulae used in these searches that we perform every day were originally created and implemented without the aid of calculators or computers. Just as the reformation is succinctly summed up in the quote "Erasmus laid the egg that Luther hatched," it could equally be said of correlation that Galton laid the foundation, but Pearson built the edifice (Porter 250). "Francis Galton invented correlation, but Karl Pearson was chiefly responsible for its development and promotion as a scientific concept of universal significance (Aldrich 364)".

Francis Galton is often remembered not as a pioneer in the field of correlation or statistics, but for his work in the field of fingerprints. He was instrumental among others, such as Faulds, Herschel, Henry and Bertillon, in justifying fingerprints as a reliable method of identification of criminals to Scotland Yard (Forrest 210,220). While Francis Galton was originally a geographer and meteorologist, it was not until later in life, when he turned his gaze toward the study of heredity, that he made his most powerful contribution (Forrest ix). This contribution proved most fruitful not only for his pupils and peers, but for generations onward.

The seeds of Galton's interest in heredity came about early in his marriage and while at Cambridge. As he rubbed shoulders with England's elite he noticed that talent could be traced throughout generations.

> I have no patience with the hypothesis occasionally expressed, and often implied, especially in tales written to teach children to be good, that babies are born pretty much alike, and that the sole agencies in creating differences between boy and boy, and man

> and man are steady application and moral effort. It is in the most unqualified manner that I object to pretensions of natural equality. The experiences of the nursery, the school, the university and of professional careers are a chain of proofs to the contrary (Forrest 89).

Galton saw inequality in the abilities of men: some had better cognitive ability such as memory capacity or mathematical reasoning (Forrest 89). Later when he married Louisa, who was unable to conceive during her life, he noticed that infertility could be seen among members of her family (Forrest 85). This caused him to speculate that her own infertility was genetic. A decade earlier Quetelet had argued that Scottish chest sizes of soldiers fell along a Gaussian curve or bell curve (developed by De Moivre in 1733), that is "the law of deviating from an average." Galton argues this can apply to other features of the human body, cognitive ability and all other genetic traits (Forrest 89, 90).

Galton dedicated himself to anthropometry no doubt being influenced by his half-cousin Charles Darwin's seminal work in the animal kingdom. Galton comments on this book that influenced his own research.

> The publication in 1859 of the *Origin of Species* by Charles Darwin made a marked epoch in my own mental development, as it did in that of human thought generally. Its effect was to demolish a multitude of dogmatic barriers by a single stroke, and to arouse a spirit of rebellion against all ancient authorities whose positive and unauthenticated statements were contradicted by modern science

(Forrest 84).

Much as Darwin compares primates to humans in skeletal structure, Galton begins to measure men in all aspects in order to correlate them along an average. He published his first work on heredity, *Hereditary Genius.* In this book he groups men and their cognitive abilities into 16 groups.

> There is a continuity of mental ability reaching from one knows not what height, and descending to one can hardly say what depth. I propose...to range men according to their natural abilities putting them into classes separated by equal degrees of merit and to show the relative number of individuals included in the several classes (Forrest 90).

The top four groups contain four fifths of the entire population that represent the average cognitive ability (Forrest 91). The groups that fall above average cognition grow smaller in population as their cognition increases because the more talented are rarer. Finally in his X group he groups those one out of a million who is labeled illustrius. The group just below illustrius are 248 per million marked as eminent (Forrest 91). He concludes that this normal distribution of cognitive ability means that you will find 50,000 idiots and imbeciles out of the 'twenty million inhabitants of England and Wales (Forrest 91). It should be pointed out that Galton makes errors in the processing of his data, but the correlation concepts behind this are sound (92).

In Galton's short ten page paper, "Co-relations and their Measurement Chiefly from Anthropometric Data," that was delivered to the Royal Society,

contains the first correlation values ever calculated. "This paper contains details of his technique for calculating the correlation coefficient and presents coefficients obtained from the measurements of 350 adult males (Forrest 197)." These coefficients are some of the first correlations ever published. The number on the left is the coefficient calculated that shows relative similarity in the measurements among these 350 males. In other words, the closer the number is to one the more similar the features are in all men. Based upon Galton's coefficient of men's knee heights and statures (0.90), he could expect future measurements to be extremely similar.

0.80 Cubit (length of forearm) and stature

0.35 Head length and stature

0.70 Middle finger and stature

0.85 Cubit and middle finger

0.45 Head breadth and head length

0.90 Knee height and stature

0.80 Knee height and cubit

Galton demonstrates in this paper that these concepts of correlation have far reaching implications for all disciplines of science (Forrest 199); he discovered a general mathematical method that can be applied to any science in order to measure similarity between data. Even though Galton seems to foresee how profound his research will impact future generations, he would be shocked to see how many disciplines today still use many of his concepts. Pearson, Galton's pupil, comments on this work:

Galton's very modest paper of ten pages from which a revolution in

> our scientific ideas has spread is in its permanent influence, perhaps, the most important of his writings. Formerly the quantitative scientist could only think in terms of causation, now he can think also in terms of correlation. This has not only enormously widened the field in which quantitative and therefore mathematical methods can be applied, but it has at the same time modified our philosophy of science and even of life itself (Forrest 197-199).

These comments of Pearson can in no way be understated. Pearson is instrumental in recognizing that Galton had established a new tool to be used in science powered by mathematics. While Galton's statistical methods were recognized as important in and of themselves, Pearson was instrumental in seeing that this method would be put to immediate use by all branches of science. Using mathematical methods, scientists could use correlation coefficients in many disciplines to be given hints (Hobson 1).

Karl Pearson's work in correlation was a life-long process starting in 1891. "He codified the mathematics of Galton's statistical idea (Porter 258)." His work is so foundational to modern statistics that he is credited with coining not a few statistical concepts such as Beta distribution, Chi-squared, the coefficient of correlation, the coefficient of variation, the histogram, homoscedastic, mode, standard deviation and sampling distribution among others (David 121,122). Anyone familiar with statistics and probability would be astounded to know that virtually two men created this entire discipline. Pearson believed that correlation was so important that it related "to all science" and would usher in a profound

change in how research is done (Porter 286). He would admit, of course, throughout his lifetime of evaluating others' correlative work, that he "found more and more situations in which correlation analysis was misleading (Aldrich 366)." Pearson is careful to note that "it is possible to obtain a significant value for a coefficient of correlation when in reality the two functions are absolutely uncorrelated (Aldrich 364)." It is this dedication to precision and his religious-like fervor that makes us owe Pearson an additional debt of gratitude. For if Pearson had been so cavalier to assume all correlations were valid and always had probative value, statistical methods could have been laughed off the stage of science forever.

Basic Correlation Examples

In an effort to understand how document correlation works, we use an extremely simple test document. The contents of this document are the familiar English pangram: The quick brown fox jumps over the lazy dog. We wish to correlate this document against a second document to quantify similarity. The contents of the second document are the following: The fox jumps over the dog. We can perform an organic correlation quite quickly on our example and conclude that both documents are extremely similar since the second document only eliminates the adjectives. However, let us step through some mathematical correlations to see their strengths and weaknesses.

We start by counting the frequency of the words in each document that gives us a simple matrix. A matrix is simply columns and rows of numbers of any size. Each column corresponds to the frequency of words in a particular document otherwise known as a document vector.

Table 1.1 Example document vectors/matrix

Word	Document Vector 1	Document Vector 2
brown	1	0
dog	1	1
fox	1	1
jumps	1	1
lazy	1	0
over	1	1
quick	1	0
the	2	2

We use these document vectors to calculate each of the correlation coefficients.

While we describe all formulae in Appendix A, I believe it is important to describe these correlations in simple terms in order to understand them. The matrix above becomes our data points that can be plotted in 2-dimensional space. These data points are what we will use to calculate similarity using the various formulae outlined in Appendix A.

Pearson

The Pearson correlation that was introduced by Karl Pearson over 100 years ago is a measure of the linear similarity of a sample data set. The following image demonstrates sample distributions of data and their respective coefficients (Pearson Coefficient). The distributions below are document vector data points plotted in 2-dimensional space.

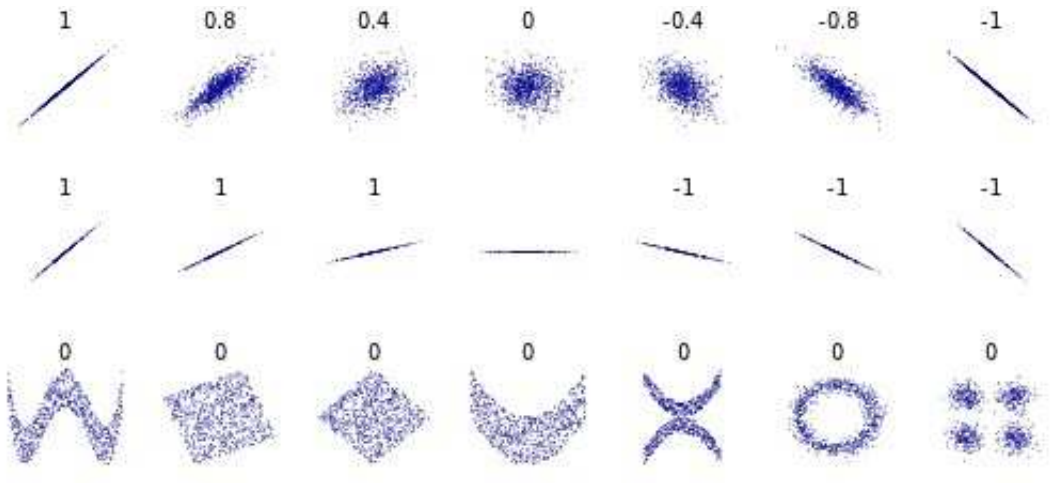


Figure 1.5 Pearson plot examples.

As you can see in order to get a positive coefficient there must exist a linear similarity with a positive slope (a slope that points up to the right).

Jaccard

The Jaccard coefficient is simply the size of the intersection of the data set divided by the size of the union of the data set. Imagine there are three document vectors represented by the three circles in figure 1.6 (Jaccard Coefficient). The intersection of the sample data is demonstrated in A of figure 1.6. The union of all three document vectors is represented by all the letters: A, B, C, D, E, F and G. So we divide the values of A by A, B, C, D, E, F and G.

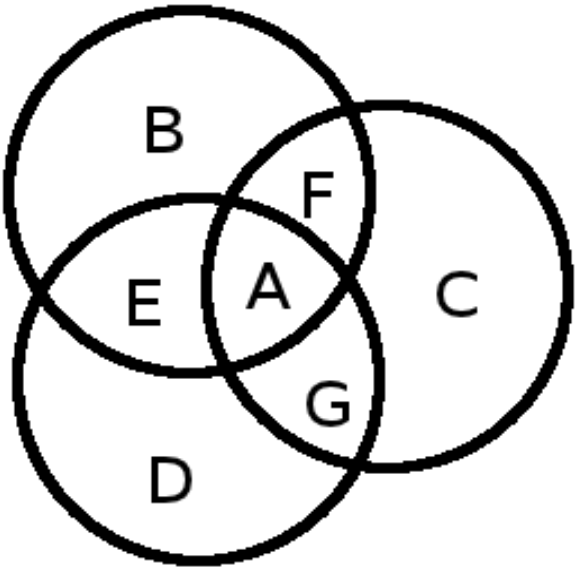


Figure 1.6 Intersection of dataset.

The Jaccard difference is simply the Jaccard coefficient minus 1, thus it tells us simply how far the Jaccard coefficient differs from a perfect similarity indicated by 1.

Cosine Similarity

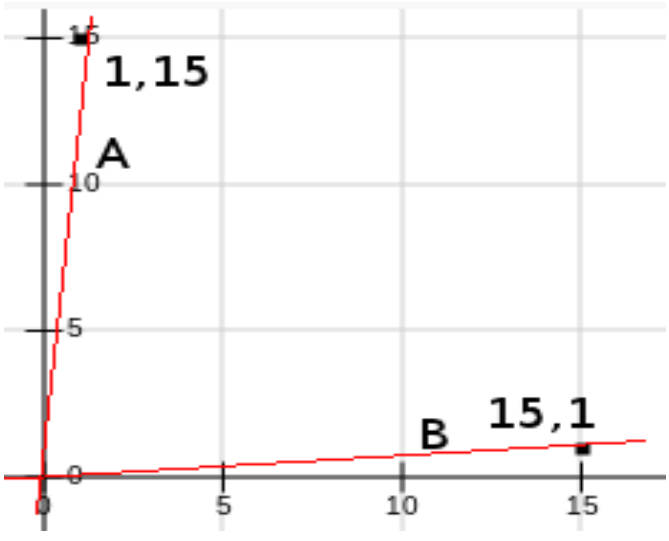


Figure 1.7 Cosine Similarity in two-dimensional space.

Figure 1.7 above is a plot in two-dimensional space. For simplicity sake we only use 2 data points. We draw imaginary lines from our data points to the origin of

our plot (0,0). We take the cosine of the angle between these lines and this is our coefficient.

Tanimoto

Many people use the Tanimoto coefficient as a synonym of the Jaccard index, but it can be mathematically distinct. This formula reduces our document vector to zeros and ones. Thus it becomes what is called a bitmap or a bit array (a list of zeros and ones). In our example a particular document vector has a zero value if a word does not appear, and a one if a word does appear. For example, if we have three occurrences of the word dog, the value in our document vector is not 3, but 1. The formula is then the number of common bits between the samples divided by a set of bits set in either sample, or all samples. Thus if you divide the common bits (the intersection A above in Figure 1.6) by A, B, C, D, E, F and G, Tanimoto could become identical to the Jaccard coefficient.

Spearman

Quite simply the Spearman coefficient is the Pearson formula with a twist. The twist is to rank the values (i.e. the frequencies of words) in ascending order and then change the respective values before performing the Pearson formula upon the new values. This process is supposed to get rid of values that are significantly larger than the rest of the sample. Spearman is then touted to be a better Pearson in certain circumstances.

Euclidean Dot Product

The Euclidean Dot Product is the sum of the products of individual entries of our document matrix. For example, if we have a document vector A (1,3,0) and a document vector B (2,0,3). These values correspond to two documents with their

respective frequencies of words. These values are then multiplied together across document vectors and added up: (1 • 2) + (3 • 0) + (0 • 3) = 2. Using this value we then can calculate the cosine of the angle between our document vectors: cosθ (where θ represents the angle) = 2 / (square root ($1^2 + 3^2 + 0^2$) * square root($2^2 + 0^2 + 3^2$)). Table 1.2 contains all of the coefficients described above for our simple pangram:

Table 1.2 Example pangram coefficient correlations.

The quick brown fox jumps over the lazy dog	The fox jumps over the dog
Pearson	0.71429
Jaccard Similarity	0.66667
Jaccard Distance	0.33333
Tanimoto Coefficient:	0.66667
Tanimoto Difference	0.33333
Spearman	0.41667
Cosine Similarity	0.85280
Euclidean Dot Product	0.45110
Euclidean Distance	0.47492

For the Pearson coefficient, the Jaccard Similarity, the Tanimoto coefficient and the Spearman coefficient a value closest to 1 indicates a theoretically perfect correlation. It is to be noted that these numbers themselves do not indicate an absolute correlation, but as Pearson said above, they are a hint (Hobson 1). In other words, the coefficients in question can all be 0.99999 and the data itself could diverge greatly. We are well-advised by Karl Pearson in his relentless skepticism of any published correlations. Compare the following table where the two documents compared differ by only 1 word.

Table 1.3 Example pangram coefficient correlations with dogs.

The quick brown fox jumps over the lazy dog	The quick brown fox jumps over the lazy dogs
Pearson	0.83205
Jaccard Similarity	0.80000
Jaccard Distance	0.20000
Tanimoto Coefficient:	0.80000
Tanimoto Difference	0.20000
Spearman	0.62424
Cosine Similarity	0.90909
Euclidean Dot Product	0.47683
Euclidean Norm	0.73242
Euclidean Distance	0.34527

We would expect a higher correlation given that the documents differ by only 1 word. Perhaps the relatively low coefficients have to do with our small documents. To illustrate this, we take the first chapter of *Moby Dick* and change the two instances of Ishmael to Israel in the second document. Our suspicion is confirmed with the new coefficients that our test documents were too sparse in data.

Table 1.4 Call me Israel coefficients.

Call me Ishmael document	Call me Israel document
Pearson	0.99996
Jaccard Similarity	0.99540
Jaccard Distance	0.00460
Tanimoto Coefficient:	2.33154
Tanimoto Difference	-1.33154
Spearman	0.99216
Euclidean Dot Product	0.56039
Euclidean Norm	0.74861

Euclidean Distance	0.00686

The Pearson coefficient, the Jaccard Similarity and the Spearman are so close to 1 that their values are almost perfect matches. This confirms that if our data is sparse, it could yield a relatively low coefficient against a very similar document. Therefore, low coefficients do not always indicate dissimilar documents. The Tanimoto Coefficient displayed in Table 1.4 is similar in its calculation to the Jaccard Similarity (aka Jaccard Index), but it is distinct as described above (See Appendix A).

Let us compare the first chapter of *Moby Dick* again to a second document containing only the first paragraph of this same chapter. For ease of understanding we calculate only the Pearson Coefficient: 0.80346. This is a relatively low correlation coefficient. We can change the data slightly to account for the differing document sizes. We do what is called normalizing the vector values by adding up all the values of the entire vector, and then divide each single value by this total. Instead of a clean matrix with whole numbers, our result is a matrix with decimals. We decide to keep 5 significant digits. Using this matrix our Pearson coefficient result is not much different: 0.80350. In both documents there are many words that are insignificant. These words that we desire to exclude are called stop words. They are words to which we always assign a zero value so that our calculation knows these words are irrelevant or too common. For example, if we were correlating two documents with the content below in Table 1.5, the result (0.85968) would be a relatively high coefficient. Many words below are inconsequential, but are being used in this

calculation. We do not want words such as demonstrative pronouns, relative pronouns, articles, etc.; otherwise they skew our coefficients.

Table 1.5 Negative correlation.

Document 1	Document 2
This is a dog, which is really a canine.	This is the pericardium, which is really a membrane.

This is a simple example, but it demonstrates how two documents could be highly correlated and differ wildly in content. In our calculation, we simply tell the Pearson algorithm to ignore all words that we deem insignificant: this, is, a, really, which and the. A decision has to be made whether we want to exclude these words all together or simply make all their values zero. (A different coefficient will result depending upon inclusion or exclusion of these zeros.) We decide to exclude them completely for clarity sake. Our document matrix looks like the following:

Table 1.6 Negative correlation document matrix.

Word	Document 1	Document 2
canine	1	0
dog	1	0
pericardium	0	1
membrane	0	1

Our coefficient in this calculation is -1.00000, a perfect negatively correlated document, thus absolutely dissimilar. Values can be negative that indicate conversely, a negatively correlated document, i.e. disimilar documents. Using our *Moby Dick* example, we create a stop word document (Appendix C) to

compare again the first paragraph to the entire first chapter to see if our result is any different. Our coefficient becomes 0.60693. While our result is different, the coefficient in question does not instill confidence in our method. Let us compare this coefficient against the first paragraph of thirty other random chapters. If our coefficient is "low," perhaps against other chapters it will seem "high." In fact, after comparing the first paragraphs of thirty random chapters (see Appendix E), there is not a single coefficient above 0.07. Our coefficient of 0.60693 becomes an extremely significant number when juxtaposed against these other coefficients. Therefore, it is evident that while results can vary wildly, the context of coefficients is critical. We also need to be mindful to eliminate within the data itself, that which is insignificant noise.

At this point in our calculations we need to start excluding correlation coefficient algorithms that do not help us compare documents accurately *for our purposes*. As has been stated, we cannot simply assume these formulae are magic and give us absolute proof as to whether our documents are truly related or not. This largely depends upon our data, i.e. the documents in question. We also cannot negate the organic element to correlation. As stated above, Pearson was mindful of this organic element: false-positives have to be assumed until we glean evidence to confirm the coefficient in question.

The Jaccard and Tanimoto coefficients are excellent similarity tools. These particular correlations are still widely used to compare chemical compounds and genes in molecular biology as well as organic chemistry. We, however, cannot use them. We exclude them because they both emphasize the presence of common features and neglect the absence of common features (Fligner,

Verducci and Blower 111; Todeschini and Consonni 699). This means that our use of these correlations depends upon what we are comparing. The absence of certain words in our documents could equally be significant to those words that are present. These correlation coefficients do not take this into account, thus we exclude them.

The coefficient that we have been using for our examples is the Pearson correlation; this is the flagship of Karl Pearson's life work. It does not suffer from the problems of the Jaccard and the Tanimoto coefficients, but it does have a known limitation. "If the data from the rating scale tend to be skewed toward one end of the distribution, this will attenuate the upper limit of the correlation coefficient that can be observed. The coefficient can appear inflated in certain circumstances (Osborne 39)." In other words, if a particular document has an unduly large frequency of a particular word, the coefficient may result in a high correlation, but in actuality indicates a high frequency of the single word in question. A simple example will illustrate these problems. We start with two documents that compare a simple sentence. The first sentence has adjectives while the second sentence excludes them. Our Pearson coefficient was 0.71429. We add the word skewed 90 times to the first document and 110 to the second document. The addition of this single word raises our coefficient to 0.99992. If we change the first document to have 10 instances of skewed and the second to have only 50, we still end up with a very high coefficient: 0.99932. The Pearson correlation does not handle these types of documents well. That is, if your documents have a few data points that are significantly larger than the rest, they will skew your results. Care must be taken then, to either eliminate

these larger numbers (outliers), or weight them differently. For now we exclude the Pearson correlation.

Table 1.7 Skewed correlation.

Word	Document 1	Document 2
quick	1	0
brown	1	0
fox	1	1
jumps	1	1
over	1	1
the	2	2
dog	1	1
lazy	1	0
skewed	90	110

Our next logical step would be to examine the Spearman correlation since it does not suffer from this problem. Spearman's Rho (Appendix A) is calculated as 0.41667 in our example without outliers and then 0.70909 with both 90/110 instances of the word skewed and with the example of 10/50 instances of the word skewed. It seems to account for these outliers and gives us a coefficient that is not too highly correlated. Spearman's rho seems like a great candidate for our purposes, but an underlying assumption is that your data has a monotonic relationship (Wikipedia). If the frequency of a given word in document 1 increases, the frequency of that same word never decreases in document 2--this is a monotonic relationship. Or stated conversely, as the frequency of a given word in document 1 increases, the frequency of that same word never increases in document 2. We could not justify such a causal relationship with our data, therefore our data is not monotonic and Spearman's Rho should not be used.

The Cosine Similarity is a powerful tool in computing document simlarity. To illustrate how this coefficient is calculated, take the following simple two-dimensional document vectors.

Table 1.8 Bad Cosine Similarity example.

Word	Document 1	Document 2
puer	1	14
puella	1	14

Here is the corresponding plot of each document vectors.

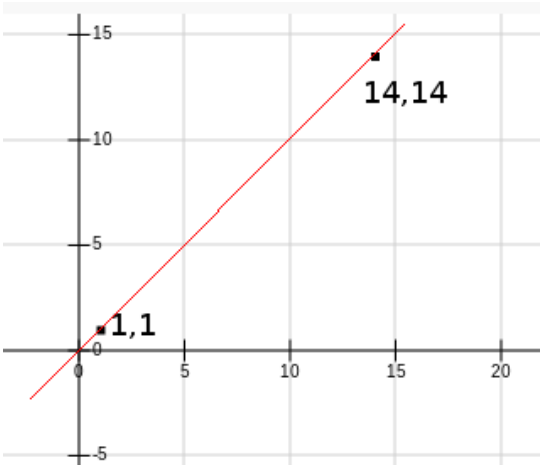


Figure 1.8 Plot of bad Cosine Similarity.

As you can see the document vectors plotted in two-dimensional space (plotted from 1,1 and 14,14 respectively) are actually right on top of each other. Normally you would measure the cosine of the angle from each of these points through the origin (0,0), but there is no angle to measure that signifies a perfectly correlated document. The Cosine similarity coefficient in our example has a value of 1, an ostensibly exact match even though in two-dimensional space they are relatively far away from one another. If document 2 had 1 reference to puella and 15 references to puer, and document 1 had 15 references to puella and 1 reference

to puer, the plot would look like figure 1.9 below.

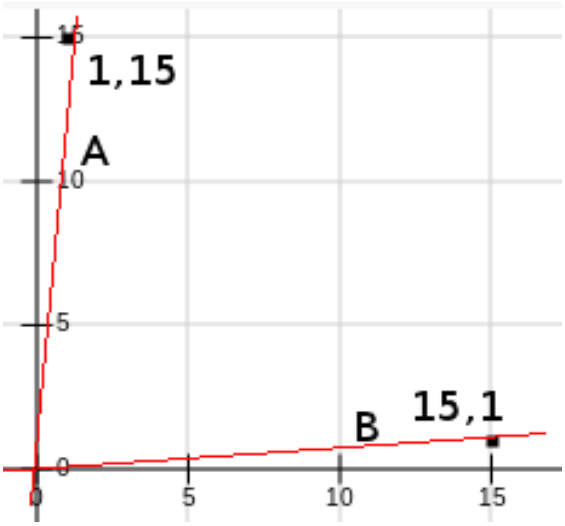


Figure 1.9 Plot of bad Cosine Similarity 2.

In such case, we measure the cosine of the angle between line A and line B through the origin (0,0) that gives us the value 0.13274. This is an extremely low coefficient, but when compared with our first example, the documents are not all that different. Both documents mention both target words, but because of their relation to each other in two-dimensional space one correlates highly, the other does not. As in our other coefficients, a value close to 1 corresponds to similarity. Now imagine the vector for document 1 is unchanged (1,1), but for document 2 we change the instances of puer to 100 (1,100). The Cosine coefficient is 0.71414. This is a drastic change by only changing the frequency of one of the words. In fact, within three examples where all documents have the same words, we have three very different coefficients. The Cosine similarity is not useless, but because it does not take into account the magnitude of the vectors (their length), it is not the wisest choice for our data. The Euclidean Norm also suffers from this problem.

We admit that we could simply run an algorithm to determine outliers and eliminate them before running a correlation, but we would much rather keep our document vectors intact. We desire a method that accounts for the entire document vector without ignoring the absence of words. It also should not assume an underlying causal relationship between the document vectors. Singular Value Decomposition may help us in our endeavor to process our matrix before running a correlation.

Singular Value Decomposition (SVD) was developed by Beltrami and Jordan in the 1870s and extended later by Golub in the 1960s (Long 161). Many scientists, mathematicians and scholars describe SVD as a way to simplify a given matrix (Long 161; Alvo and Ertas 482; Good 823; Hubert, Meulman and Heiser 69). This simplification exposes the underlying geometric structure, that allows us to understand better the way the vectors relate to each other. It has been used over the past 50 years for a variety of applications. It has been used to correlate areas of the brain (Worsley 915), to classify or organize genes in organic chemistry (Yeung 6163), to summarize differences in solar radiation that vary by geographical location (Glasbey 382), image processing (Long 164-166), to relate genes within DNA studies (Omberg and Golub 18731), to assist in screening certain patients for different cancer treatments (2052) and in text processing (Alvo and Ertas 482; Alvarez-Lacalle, Dorow and Eckmann 7956-7959). While it appears to be perplexing to the classicist that an algorithm that has been so prevalent in scientific studies, can be used to correlate texts, it is completely natural since we can reduce our texts to a column of numbers, i.e. a document vector as seen above. Some have indicated additionally that if we can

represent accurately SVD this may help not only in our efforts to communicate the importance of this algorithm to others, but Hubert, Meulman and Heiser continue that this representation may also further our own understanding of our own data (69). While this is certainly interesting, representing our data spatially is beyond the scope of this dissertation. We shall be content to use SVD and calculate coefficients after we have run this algorithm on our matrices.

SVD can be thought of as a simplification as stated above, or a factorization. We can factor the number 66 that results in 11 and 6 because 6 multiplied by 11 = 66. Instead of starting with a whole number we start with document vectors or a matrix. The following columns can be thought of separately as individual document vectors, or as a complete matrix.

Table 1.9 SVD simple example - unprocessed document matrix.

Matrix A			
Words	**Document 1**	**Document 2**	**Document 3**
when	1	1	3
the	1	2	3
drops	1	0	4
start	1	0	1
stopping	1	1	0
the	1	2	0
rain	1	0	0
starts	1	1	2
stopping	1	2	2

It is this 3 x 9 matrix that we could decompose or factor using SVD. It is thought that this process exposes underlying properties of the matrix, that would otherwise be unrealized. These properties have to do with the geometric

structure of the matrix. This matrix is decomposed or factored into three component parts (three matrices) in Singular Value Decomposition (SVD). These are denoted as Σ V and U, and if multiplied together will give us our original matrix. Σ is a rectangular diagonal matrix (a diagonal matrix is one where the values outside the main diagonal are zeros, see Table 1.10 below) where the values are not negative. U is an orthogonal matrix, and V is another orthogonal matrix that has been transposed. SVD has a "unique mathematical feature of providing the rank-k approximation to a matrix A of minimal change for any value of k (Berry 53)." This means that a given matrix A, when decomposed with SVD, will give us special values in Σ. Σ is a matrix of singular values that we can choose to use or eliminate. Our example below has only three values (7.47941, 3.02687 and 1.37712), but we could easily have a matrix of many values in another example. From this matrix, we could choose any number of values to calculate our new matrix. A different matrix results depending upon how many values we choose. Our matrix then becomes a rank-5 approximation if we choose 5 values, or a rank-4 approximation if we choose 4 values and so on. For example, in our matrix A above, when factored, we get the following matrices.

Table 1.10 SVD simple example - Σ.

Σ		
7.47941	0.00000	0.00000
0.00000	3.02687	0.00000
0.00000	0.00000	1.37712

Table 1.11 SVD simple example - U.

U		
0.44069	0.12071	0.03309
0.49270	-0.15135	0.33300
0.50277	0.55499	-0.13945
0.16051	0.06833	-0.52157
0.09843	-0.36595	-0.34903
0.15044	-0.63801	-0.04912
0.04643	-0.09389	-0.64894
0.32661	-0.04151	-0.09429
0.37861	-0.31357	0.20563

Table 1.12 SVD simple example - V^T.

V^T (V^T is a matrix V transposed, see Appendix A)		
0.34725	0.38896	0.85330
-0.28420	-0.82349	0.49102
-0.89367	0.41301	0.17541

From Σ above, we can chose only to use 2 singular values 7.47941 and 3.02687. When we multiply ΣUV^T, using only 2 singular values, this results in a rank-2 approximation of our original matrix as discussed above. This new matrix is mathematically similar to our original matrix, but in certain cases can reveal similarities in document vectors. Currently there is no automatic method that reveals the optimal rank to choose. Ranks are chosen by empirical testing (Berry 54).

Granted, SVD does not give us coefficients between document vectors, but it does factor our matrix based upon all document frequencies. Thus, as we examine SVD it does not suffer from the same problems as some of our other

correlation types. Unlike the Tanimoto and Jaccard index, SVD does take into account the absence of common features. Additionally SVD does not suffer from the flaw of outliers as in the Pearson correlation. SVD also does not assume a monotonic relationship like Spearman's Rho. While SVD is not a silver bullet in and of itself, it will give us a good base from which to run our coefficients. After we recalculate a given matrix using SVD we then use a specific correlation algorithm to compare the document vectors of our new approximated matrix (below we will choose Pearson's correlation after SVD). In simple terms, SVD fixes the document vectors in our matrix.

Let us prove this with a simple example. We start with a similar example as before.

Table 1.13 SVD example - simple pangram document matrix.

Word	Document 1	Document 2
quick	1	0
brown	1	0
fox	1	1
jumps	1	1
over	1	1
the	2	2
dog	1	1
lazy	1	0

After processing our matrix through SVD we obtain the following new matrix.

Table 1.14 SVD example - simple pangram after SVD.

Word	Document 1	Document 2
quick	0.59215	0.49144
brown	0.59215	0.49144
fox	1.08358	0.89929
jumps	1.08358	0.89929
over	1.08358	0.89929
the	2.16716	1.79858
dog	1.08358	0.89929
lazy	0.59215	0.49144

Notice that values in our document vectors that were previously zero are now above 0.0. SVD factored our matrix and processed it geometrically to derive at different document vectors, but that are related to each other. We now use Pearson coefficient against the document vectors and receive a 1.00000 correlation coefficient. Previously we excluded Pearson because of outliers, but since our matrix has been processed to eliminate outliers we feel safe using it. We rightly receive a perfect correlation coefficient since our documents only differ in a few words. When we again add our outliers our coefficient does not change. It is again, 1.00000 (with the outliers of 90 and 110). Thus we are confident in our method to compare documents.

Chapter 2 - Our Method

I used texts at TheLatinLibrary.com and Perseus.org and programmatically separated them into their component books and poems. For Lucilius no reliable online text was found so I typed up the fragments based upon the Loeb edition and entered all of these texts into MySQL, an open-source database. It is from this database that I performed all operations. In addition to the poems and fragments, it was necessary to import into my database a Latin dictionary that I obtained from the *Perseus Project* website in order to extract lemma forms for words or to indicate tenses for any tense correlations. For correlations using proper nouns, I went through all the satirists and flagged these nouns in the database (see Appendix B). For the special subject correlations I created 11 categories (see Appendix B) based upon known satire themes: animals, disease, excess, food, man and virtue, speech, the body, the dishonorable, the gods, war language and women. From the fragments of Lucilius I then imported all words that correspond to these categories.

In order to perform correlations on the target documents, whether they are entire books or single poems of satire, these texts are extracted from the database. These words are extracted depending upon the type of correlation we are doing, e.g. lemma, proper nouns, exact words, subject words, etc. We also excluded common words using the stop words (see Appendix C) mentioned above so that document similarity is not skewed by words like simple

conjunctions, pronouns, etc. A unique list of all words across the target documents is created and a document vector (column of numbers) is populated with each document's word frequencies. Thus if a word appears in Horace 12 times, the value will be 12, or if a particular word is not used at all, a zero is used. This simple matrix is not normalized. Normalization means that an algorithm changes a given matrix slightly to account for relative document lengths. One such normalization technique is to add up all the squared word frequences of the entire vector. We then take the square root of that value and then divide each single value by this new value. For example, we have a document vector A [1,3,0]. We could normalize this document vector: square root($(1^2 + 3^2 + 0^{2)}$)=3.16227. We take each value of our document vector A and divide by this new value: 1 / 3.16227, 3 / 3.16227, 0 / 3.16227. We receive a normalized document vector A [.31611, .94868, 0]. To clarify, we do not perform normalization. This matrix is processed using the Singular Value Decomposition algorithm with a rank-k approximation (e.g. 4 singular values could be used to create our new decomposed matrix, the tool described in chapter 7 can be changed to use any number of ranks). At this point, we have a more accurate representation of document similarity because of our factoring. This means, that theoretically, a previously zero value denoting word frequency in a document matrix can be incremented because the document in question has other values that indicate to SVD a particular frequency needs to be higher than it actually appears in the original document vector. We saw this above with our simple example. SVD can therefore change a given document matrix.

In our correlations, we can also use multiple words instead of single

words. For example, in this line of poetry, "once upon a midnight dreary, while I pondered, weak and weary," we could create a document vector with 11 words, as in Table 2.1.

Table 2.1 Poe document matrix.

Word	Count
once	1
upon	1
a	1
midnight	1
dreary	1
while	1
I	1
pondered	1
weak	1
and	1
weary	1

We could also choose an index of 3 words. Therefore, we would have a matrix that would look like Table 2.2.

Table 2.2 Poe document matrix index=3.

Phrase	Count
once upon a	1
upon a midnight	1
a midnight dreary	1
midnight dreary while	1
dreary while I	1
while I pondered	1
I pondered weak	1
pondered weak and	1
weak and weary	1

As you can see, these document vectors are quite different and could yield much different results. We can change this index to suit our correlative needs. It may however, be less useful when we are dealing with an author like Lucilius, who exists in fragments.

We use this same method (SVD and then Pearson to measure document vectors) when comparing individual poems except that our document vectors are much shorter since the content from our target documents are shorter. In like manner, when comparing the fragments of Lucilius, these document vectors are even smaller, and will perhaps be less accurate depending upon fragment lengths. Therefore we will need to alter our method slightly when comparing these fragments because of the paucity of words in each Lucilian fragment. We must then compare separate poems to the books of Lucilius instead of individual fragments.

Last, for the unassigned fragments and other poems I will perform what I term a roving correlation against the books of Lucilius. I will take a particular fragment and count the lines of the fragment. I will then run a correlation with that fragment against individual fragments of the individual books of Lucilius (I-XXX). For example, I will take line 1221 that consists of 4 Latin words. I will correlate this against book I of Lucilius, line 1; and then against Book I, line 2; and then against Book I, lines 3 and 4 (because lines 3 and 4 are a single fragment); and so on. In this way, data will be generated to indicate if a particular unassigned fragment correlates highly to a particular fragment within a book of Lucilius. Fragments of Lucilius that have only a few usable words (those that are not stop words) were not good candidates for us.

To mitigate correlations against inconsequential or common words, we use the classical stop words that Perseus uses in their Lucene/Solr document search (see Appendix C). After we calculate the SVD of a given matrix we then take the Pearson correlation coefficient against each document vector.

For a simple sanity check of our method, we select a test that no secular or biblical scholar would dispute. There are 27 books of the New Testament, thirteen of which present the Apostle Paul as their author. Even if someone were to claim some of these books were written by another author, no one would dispute that these epistles claim to have one author and have marked similar language when compared with the other books of the New Testament. Therefore we should see high coefficients when we compare these books using our method. Additionally, other books of the New Testament should have relatively lower coefficients since they have different content, e.g. the Gospels. We could have done these coefficients against the Greek New Testament, but since we will be shortly running coefficients on Latin works we thought it best to use the Vulgate.

Table 2.3 Pauline coefficient correlations using Galatians.

Book	Number of Words	Correlation Coefficient
Galatians	1172	1.00000
Ephesians	1307	0.95851
Philippians	934	0.99121
Colossians	857	0.98359
1 Corinthians	3759	0.99419
2 Corinthians	2500	0.99279
Romans	3780	0.98317

Noncanonical Letter to the Laodiceans		
Epistle to the Laodiceans	151	0.98994
Some Non-Pauline Epistles		
1 Peter	983	0.96166
2 Peter	660	0.97843
James	1006	0.97333
Hebrews	2872	0.96970
Gospels & Acts		
Matthew	10278	0.80828
Mark	6388	0.72321
Luke	11230	0.79565
John	8515	0.86046
Acts	10201	0.90627

We do in fact see exactly what is expected. We took the book of Galatians and correlated it against a few books of the Latin New Testament. Notice that the gospels do not have high coefficients at all. The book of Acts seems to be the most highly correlated of that set perhaps because the content of Acts describes the work of Paul, and perhaps contains similar language. The Pauline epistles contain familiar language and therefore almost all of them have high coefficients. I included the noncanonical epistle to the Laodiceans. This letter is purportedly written by the Apostle Paul, but was never considered canonical by either Protestants or Catholics. It is ostensibly mentioned in Colossians 4:16, "And when the letter is read aloud to you, take care that also it may be read aloud to the church at Laodicea; and also you should read aloud the letter coming from Laodicea." This letter's coefficient tells us that it contains much of the same language as Paul's letters and therefore if it is not genuine, the person who wrote it imitated Paul's vocabulary well. The book of Hebrews is relatively low. This

could be used as fuel to the age-old debate whether or not Paul wrote it. We try this again using the book of Philippians.

Table 2.4 Pauline coefficient correlations using Philippians.

Book	Number of Words	Correlation Coefficient
Philippians	934	1.00000
Galatians	1172	0.99121
Ephesians	1307	0.97312
Colossians	857	0.99471
1 Corinthians	3759	0.99246
2 Corinthians	2500	0.99893
Romans	3780	0.99126
Noncanonical Letter to the Laodiceans		
Epistle to the Laodiceans	151	0.99591
Some Non-Pauline Epistles		
1 Peter	983	0.98641
2 Peter	660	0.96503
James	1006	0.97161
Hebrews	2872	0.95505
Gospels & Acts		
Matthew	10278	0.75527
Mark	6388	0.66497
Luke	11230	0.75717
John	8515	0.80110
Acts	10201	0.86152

Again, we see the same stark contrast between the books of the Latin New Testament. First Peter is a little higher than the correlations that were run previously (+.02475), but the book of James (+.00172) and Hebrews (+.01465) are almost exactly the same values. The Gospels are even lower in coefficients, and again the epistle to the Laodiceans is amazingly high.

As seen by our method, the books that we know to be highly correlated correlate as predicted.

Using our Method

Confident in our method we turn entirely to Roman Satire. We now compare the books alone against one another to see how they correlate. If our specific method were susceptible to skewing coefficients based upon document lengths we would expect Juvenal to always correlate the highest to Lucilius, because he uses the most words. Compare the following table for the number of words of each satirist.

Table 2.5 Roman Satire corpus correlations.

	All Words	Words Correlated	Lucilius	Horace	Persius	Juvenal
Lucilius	7,623	5066	1.00000	**0.86359**	0.74029	0.79181
Horace	14,278	10691	**0.86359**	1.00000	0.79457	0.80912
Persius	4,521	3145	0.74029	0.79457	1.00000	0.74134
Juvenal	24,436	17365	0.79181	0.80912	0.74134	1.00000
Total Unique Words		18211	**Total Unique Correlated Words**			14257

As can be seen from Table 2.5 Horace is the highest correlated author against Lucilius. While this coefficient is not above 0.90, it is still highly signficant when compared to Persius and Juvenal. There is a difference between the exact words of each author and the words we use to correlate in our document matrices because of stop words, that are excluded, and also words for which we do not have lemma information in our database. Thus, this is a comparison based upon lemma words alone. Let us run some coefficients against exact words to see if there is a difference in coefficients. We run correlations using

specific subject correlations, proper nouns and various indices to see if another author rises to the top in each instance. You can find the list of the words used to do the subject correlations as well as the proper names correlation below in Appendix B.

Table 2.6 Subject correlations: Literal words.

Literal Words	All Words	Words Correlated	Lucilius	Horace	Persius	Juvenal
Lucilius	7,623	6223	1.00000	**0.75510**	0.52035	0.59736
Horace	14,278	11691	0.75510	1.00000	0.64022	0.68710
Persius	4,521	3706	0.52035	0.64022	1.00000	0.62466
Juvenal	24,436	19952	0.59736	0.68710	0.62466	1.00000

Table 2.7 Subject correlations: Proper names.

Proper Names	All Words	Words Correlated	Lucilius	Horace	Persius	Juvenal
Lucilius	7,623	296	1.00000	-0.08542	**-0.05083**	-0.28699
Horace	14,278	575	-0.08542	1.00000	-0.02858	-0.18644
Persius	4,521	158	-0.05083	-0.02858	1.00000	-0.11355
Juvenal	24,436	1269	-0.28699	-0.18644	-0.11355	1.00000

Table 2.8 Subject correlations: Animals.

Animals	All Words	Words Correlated	Lucilius	Horace	Persius	Juvenal
Lucilius	7,623	51	1.00000	**0.38586**	0.15617	0.23856
Horace	14,278	44	0.38586	1.00000	0.37696	0.51616
Persius	4,521	24	0.15617	0.37696	1.00000	0.40370
Juvenal	24,436	66	0.23856	0.51616	0.40370	1.00000

Table 2.9 Subject correlations: Disease.

Disease	All Words	Words Correlated	Lucilius	Horace	Persius	Juvenal
Lucilius	7,623	68	1.00000	0.46942	0.44893	**0.59994**
Horace	14,278	131	0.46942	1.00000	0.45088	0.38371
Persius	4,521	33	0.44893	0.45088	1.00000	0.72166
Juvenal	24,436	168	0.59994	0.38371	0.72166	1.00000

Table 2.10 Subject correlations: Excess.

Excess	All Words	Words Correlated	Lucilius	Horace	Persius	Juvenal
Lucilius	7,623	219	1.00000	0.79447	0.57873	**0.83576**
Horace	14,278	268	0.79447	1.00000	0.63423	0.97177
Persius	4,521	70	0.57873	0.63423	1.00000	0.60346
Juvenal	24,436	492	0.83576	0.97177	0.60346	1.00000

Table 2.11 Subject correlations: Food.

Food	All Words	Words Correlated	Lucilius	Horace	Persius	Juvenal
Lucilius	7,623	129	1.00000	**0.65563**	0.15643	0.60047
Horace	14,278	181	0.65563	1.00000	0.27231	0.67037
Persius	4,521	58	0.15643	0.27231	1.00000	0.26468
Juvenal	24,436	272	0.60047	0.67037	0.26468	1.00000

Table 2.12 Subject correlations: Speech.

Speech	All Words	Words Correlated	Lucilius	Horace	Persius	Juvenal
Lucilius	7,623	126	1.00000	0.87829	0.89965	**0.93607**
Horace	14,278	247	0.87829	1.00000	0.88154	0.92289
Persius	4,521	85	0.89965	0.88154	1.00000	0.94773
Juvenal	24,436	254	0.93607	0.92289	0.94773	1.00000

Table 2.13 Subject correlations: The body.

The Body	All Words	Words Correlated	Lucilius	Horace	Persius	Juvenal
Lucilius	7,623	169	1.00000	0.67225	0.46291	**0.68165**
Horace	14,278	178	0.67225	1.00000	0.60772	0.71015
Persius	4,521	117	0.46291	0.60772	1.00000	0.46736
Juvenal	24,436	380	0.68165	0.71015	0.46736	1.00000

Table 2.14 Subject correlations: The dishonorable.

The Dishonorable	All Words	Words Correlated	Lucilius	Horace	Persius	Juvenal
Lucilius	7,623	170	1.00000	**0.82135**	0.58623	0.68318
Horace	14,278	353	0.82135	1.00000	0.63436	0.72212
Persius	4,521	57	0.58623	0.63436	1.00000	0.71621
Juvenal	24,436	405	0.68318	0.72212	0.71621	1.00000

Table 2.15 Subject correlations: The gods.

The gods	All Words	Words Correlated	Lucilius	Horace	Persius	Juvenal
Lucilius	7,623	42	1.00000	**0.86505**	0.43647	0.45716
Horace	14,278	100	0.86505	1.00000	0.66143	0.52745
Persius	4,521	33	0.43647	0.66143	1.00000	0.60959
Juvenal	24,436	197	0.45716	0.52745	0.60959	1.00000

Table 2.16 Subject correlations: War language.

War Language	All Words	Words Correlated	Lucilius	Horace	Persius	Juvenal
Lucilius	7,623	111	1.00000	**0.75036**	0.31759	0.63732
Horace	14,278	127	0.75036	1.00000	0.38536	0.80422
Persius	4,521	26	0.31759	0.38536	1.00000	0.35830
Juvenal	24,436	225	0.63732	0.80422	0.35830	1.00000

Table 2.17 Subject correlations: Women.

Women	All Words	Words Correlated	Lucilius	Horace	Persius	Juvenal
Lucilius	7,623	63	1.00000	**0.87409**	0.75150	0.73127
Horace	14,278	126	0.87409	1.00000	0.59468	0.72836
Persius	4,521	23	0.75150	0.59468	1.00000	0.46092
Juvenal	24,436	223	0.73127	0.72836	0.46092	1.00000

Table 2.18 Subject correlations: Man & virtue.

Man & Virtue	All Words	Words Correlated	Lucilius	Horace	Persius	Juvenal
Lucilius	7,623	339	1.00000	**0.61653**	0.53585	0.60171
Horace	14,278	516	0.61653	1.00000	0.73667	0.77336
Persius	4,521	101	0.53585	0.73667	1.00000	0.72383
Juvenal	24,436	682	0.60171	0.77336	0.72383	1.00000

Out of the thirteen correlations above, Horace correlates the highest to Lucilius eight times. The proper name correlation gives us negatively correlated values indicating conversely the lack of any significant correlation. The literal word correlation, just like the lemma word correlation shows Horace to be the highest correlated author to Lucilius.

Horace is the Highest Correlated Author to Lucilius

Our method confirms what antiquity has first borne out, and subsequently what scholars have demonstrated, i.e. Horace is the highest correlated to Lucilius.

> Satura quidem tota nostra est, in qua primus insignem laudem adeptus Lucilius quosdam ita deditos sibi adhuc habet amatores ut eum non eiusdem modo operis auctoribus sed omnibus poetis

> praeferre non dubitent. Ego quantum ab illis, tantum ab Horatio dissentio, qui Lucilium 'fluere lutulentum' et esse aliquid quod tollere possis putat. Nam et eruditio in eo mira et libertas atque inde acerbitas et abunde salis. Multum est tersior ac purus magis Horatius et, nisi labor eius amore, praecipuus. Multum et verae gloriae quamvis uno libro Persius meruit. Sunt clari hodieque et qui olim nominabuntur (Quintilian 10.1.93).
>
> Satire indeed is entirely ours, in which Lucilius was the first one who obtained notable praise. Thus he still has some devotees given over to him that do not hesitate to prefer him not only above authors of similar works, but even over all poets. I disagree as much as with them as with Horace who thinks Lucilius 'flows a muddy [stream]' and there is something that you may be able to take out. For there is in Lucilius both a wonderful erudition and freedom, that makes for a biting and ample wit. By far, Horace is more polished and pure, unless I err concerning him being particular to him because of my love for him. Persius has also gained a great deal of true praise although he only has one book. There are also today some men who in the future will be called famous (Translation mine).

Quintilian admits freely that the progenitor of Roman Satire is Lucilius, but he quotes from only one author in the entire genre of Satire--Horace. Quintilian tells us that Lucilius still had devoted readers in Quintilian's time. These readers not

only preferred Lucilius over other Satirists, but preferred Lucilius above all other genres. Quintilian praises Lucilius, and takes issue with Horace for his critique of Lucilius. Horace tells us that Lucilius' poems are not as polished as they could have been. Horace jokes in 1.4 that Lucilius dictated his poems standing on one foot, or composed two-hundred lines in one hour. He says Lucilius composed his poems too quickly and needed to do the hard work of making each line as pure as possible. Quintilian appears to dissent. He states that as much as you might be tempted to see Lucilius' verses as muddy or too free, there is as much an erudition in his verses as a freedom that contributes to the whole. He states that Horace is incorrect in his assessment of Lucilius. Although he disagrees with Horace, Quintilian believes Horace to be the best of all Satirists. He even states that he could be incorrect in this assessment because of his great love for Horace. Perhaps it is Horace's estimation of Lucilius that gives us a hint why he correlates so closely with satire's progenitor. If Horace believed Lucilius to be muddied with extra things that we desire to remove, perhaps Horace wants us to think of his poems as a purer version of Lucilius. He is not so much a more polished or terse version, but a more precise version of Roman Satire.

Additionally, Quintilian mentions not only Persius and his first book, but also other writers in this genre (Donald Russell tells us this cannot be Juvenal since he wrote only after Domitian died in 96 CE and Quintilian wrote his *Institutes* previously, 303). It is significant that Quintilian mentions Persius by name. We see this significance in our correlations for Persius' first book (See Appendix D). It is highly correlated in many cases (cf. Lucilius Book 1, 26, 28, 29, 30). While this is significant, he classifies Persius with these other men who are not in the

same class as Horace. Horace seems to have a higher status than Lucilius, the progenitor of satire, and by consequence logically, the most highly correlated.

Secondary literature as well sees Horace as the most correlated to Lucilius. In Miller's anthology on 1.4 he says that even though there is a departure in 1.4 from Lucilius, "Horace both explicitly embraces Lucilius and takes his distance from his great forebearer (Miller, *Latin Verse Satire* 127)." Miller and Freudenburg both see Horace differentiating himself from Lucilius and the Old Comic roots that he mentions in verses 1-5. Others as well see distancing in 1.4, just as Anderson appropriates Juvenal's hatred and contempt as not indicative of true feelings (Anderson, "The Programs of Juvenal's Later Books" 145, 147). And additionally, Kiernan suggests, "the more objectionable or violent indignation, the more cause for separation from the poet and the persona (368)." Miller sees that the "personal, the political and the generic are so presented in this poem as to form a seamless whole (Miller, *Latin Verse Satire* 127)." This seamless whole, or farrago, that is displayed is exactly like his inventor. This could be a clue why 1.4 is so highly correlated, i.e. the topics vary because the references vary. There is language in 1.4 borrowed from many Lucilian books.

Much secondary literature not only mentions Lucilius and Horace together, but mentions specifically 1.4 as being highly Lucilian. Frank offers an interesting conjecture on 1.4. He says that as Cato thought meanly of Horace, but praised Lucilius, in 1.4.90, the *tibi* is actually Cato. That is, "Lucilius is thought to be urbane and affable to you, Cato (Frank 72)." Thus, he continues, the man pictured in the verses previous to this that we saw are so dense with Lucilian language, is Lucilius himself (Frank 73). Frank does admit some distancing of

himself within 1.4 from Lucilius, (and in 1.10, also very highly correlated), but not to the extent that he is unsympathetic. He "begins with a defense of himself against an unfair comparison with Lucilius, but sees rather important political ramifications in both (75)."

Within the first two pages of Keane's monograph on the program and genre of Roman Satire, both Lucilius is mentioned as the inventor of the genre, and within the same sentence Horace 1.4 is cited (4). Later she shows how Lucilius and Horace both demonstrate "sermoni propiora," (using plain language; 1.4.64,65; Keane, *Figuring Genre in Roman Satire* 77). She additionally says that, as she mentioned previously the invectives of Lucilius are feared, she demonstrates Horace's satire is feared as well (1.4.33,70; Keane, *Figuring Genre in Roman Satire* 78).

Hooley says that Horace 1.10 closes out the programmatic ideas first started in 1.4 (32). He correlates 1.4, 1.10 and 2.1 (all highly correlated) to Lucilius and says of 1.4, it "broaches central ideas which others (poems) will modify (Hooley 46)." Similarly, Fiske mentions all three of these poems and likens this sermo style to Lucilius as Horace "followed in the spirit of Lucilian satire (Fiske 278)." He further adds that 1.4 is an "allusion to the conscious feeling of Lucilius (279)."

> "The 4th satire may be regarded as an aesthetic and ethical analysis of the Lucilian theory of satire; a criticism, however, presented under the guise of an attack upon contemporaries who believed in a direct revival of the Lucilian invective presented in the traditional Lucilian form of improvisation (Fiske 279)."

So while, there is criticism of Lucilius, Horace does not engage in an impromptu poem prone to clumsy language, long-windedness and muddy thoughts. This is a bold reference made to this inventor of Horace's genre.

Gowers says Lucilius is autobiographical just as Horace's satires. Horace in 1.4 gives an illusion of authenticity as does Lucilius, that were both simply masks, or personas ("Fragments of Autobiography in Horace Satires I." 55; 75,76). The abrupt ending of Horace in 1.5 is reminiscent of Lucilius (Gowers, "The Loaded Table" 81).

While Kemp's main argument, whether we should take Horace's view on his literary program at face value or not, is far from our exemplifying secondary literature correlating Horace 1.4 to Lucilius, he does, however, make many references to 1.4 and correlates these references directly to Lucilius (63ff). Horace is in "1.4 defending the genuine satirist and therefore Lucilius as well as himself (Kemp 63)." He additionally sees a motif of morality in both Horace and Lucilius in 1.4 (Kemp 64).

Schlegel sees a strong parallel in Horace 1.4 and 1.6. As the comic poets taught Lucilius to look at vice, so did Horace's father teach his son (95). Horace conflates style with ethos. In 1.4.65 the question is asked whether this poetry is to be mistrusted. Horace answers the question by describing who he is in the rest of the poem (Schlegel 94). Lucilius and Maecenas are fathers of sorts to Horace; therefore, as he defends the genre, he also indirectly defends Lucilius, thus referencing his corpus.

Different from most of the secondary literature, direct references to

Lucilius' fragments are made by Freudenburg, correlating specific lines of Horace with those of Lucilius. He relates 1.4.88-89, some of our densely Lucilian verses with Lucilius lines 670-1, that is from book XXVI (Freudenburg, *Satires of Rome* 39).

Horace 89 condita cum verax aperit praecordia Liber:

Lucilius 670 Ego ubi quem ex praecordiis

Lucilius 671 ecfero versum,

Horace 89 when the truthful Bacchus uncovers the seasoned heart

Lucilius 670,671 When I bring forth / a line from my heart

The same type of language is seen as both of their hearts are laid bare. In Horace, the context as Frank conjectured (72) could very well be Lucilius himself, thus making the reference specific. Freudenburg writes, "Thus, Lucilius' project, as Horace constructs it in *Satire* 1.4, is an exact mirror image of the poet's swaggering, late-republican elite-male self: politically engaged, hyper-confident, unchecked, not niggling over details, prolific (Freudenburg, *Satires of Rome* 49,50)." Freudenburg sees Horace as quite Lucilian. He imitates Lucilius to a point, for he is stifled by his status. While both Horace and Lucilius enjoy the necessary libertas (freedom) to engage in Roman satire, they do not enjoy the same quality of it. Lucilius was greater in wealth and status through his well-connected family while Horace was the son of a freedman. This difference in libertas meant that Horace could imitate Lucilius, but would never sound quite precisely like him (Freudenburg, *Satires of Rome* 49-51).

Chapter 3 - Horace's highest correlated poem to Lucilius

We have confirmed technologically that Horace is the highest correlated author to Lucilius. If we can confirm the author who has been seen by scholars as the most correlated to Lucilius, we should be able to confirm specific poems that correlate highly to Lucilius as well. This becomes a little tricky since we cannot separate Lucilius with perfect confidence into specific poems. In our database we have 866 distinct fragments or sections in Lucilius. 314 of these fragments have under 7 words. Additionally, Lucilius can be broken down into 30 books. Correlations for every individual poem of Horace, Persius and Juvenal to the separate books of Lucilius are shown in Appendix D. Comparing all of Lucilius' fragments to individual poems yielded nothing significant irrespective of document length because of the volume and diversity of Roman satire itself. We separated Lucilius into separate books showing five of the these correlations below:

Table 3.1 Book 26 against individual poems.

Book 26 632 - 736		
Poems	**Poem Length**	**Coefficient**
Lucilius - Satires	498	1.00000
Juvenal - Satires - 1	770	0.80355
Juvenal - Satires - 2	737	0.86122
Juvenal - Satires - 3	1419	0.87189
Juvenal - Satires - 4	671	0.85663
Juvenal - Satires - 5	752	0.96277

Juvenal - Satires - 6	3085	0.72035
Juvenal - Satires - 7	1096	0.88927
Juvenal - Satires - 8	1169	0.94893
Juvenal - Satires - 9	696	0.98007
Juvenal - Satires - 10	1689	0.67696
Juvenal - Satires - 11	946	0.82669
Juvenal - Satires - 12	571	0.77813
Juvenal - Satires - 13	1162	0.82902
Juvenal - Satires - 14	1524	0.86182
Juvenal - Satires - 15	811	0.72220
Juvenal - Satires - 16	267	0.93164
Persius - Satires - Prologus	46	0.89257
Persius - Satires - 1	619	0.97894
Persius - Satires - 2	369	0.92163
Persius - Satires - 3	573	0.97771
Persius - Satires - 4	234	0.96666
Persius - Satires - 5	923	0.97532
Persius - Satires - 6	381	0.97785
Horace - Satires - 1.1	631	0.98893
Horace - Satires - 1.2	694	0.98730
Horace - Satires - 1.3	706	0.99623
Horace - Satires - 1.4	735	0.99883
Horace - Satires - 1.5	499	0.95469
Horace - Satires - 1.6	679	0.95293
Horace - Satires - 1.7	164	0.95888
Horace - Satires - 1.8	239	0.93349
Horace - Satires - 1.9	414	0.98278
Horace - Satires - 1.10	477	0.99947
Horace - Satires - 2.1	421	0.99008
Horace - Satires - 2.2	694	0.99929
Horace - Satires - 2.3	1657	0.85073
Horace - Satires - 2.4	448	0.98614
Horace - Satires - 2.5	568	0.99175
Horace - Satires - 2.6	612	0.99925
Horace - Satires - 2.7	592	0.97185
Horace - Satires - 2.8	461	0.98838

Table 3.2 Book 27 against individual poems.

Book 27 737 - 792		
Poems	**Poem Length**	**Coefficient**
Lucilius - Satires	278	1.00000
Juvenal - Satires - 1	770	0.82303
Juvenal - Satires - 2	737	0.87186
Juvenal - Satires - 3	1419	0.90481
Juvenal - Satires - 4	671	0.88656
Juvenal - Satires - 5	752	0.96477
Juvenal - Satires - 6	3085	0.70919
Juvenal - Satires - 7	1096	0.90482
Juvenal - Satires - 8	1169	0.96172
Juvenal - Satires - 9	696	0.97422
Juvenal - Satires - 10	1689	0.71405
Juvenal - Satires - 11	946	0.85780
Juvenal - Satires - 12	571	0.80792
Juvenal - Satires - 13	1162	0.85901
Juvenal - Satires - 14	1524	0.88889
Juvenal - Satires - 15	811	0.75424
Juvenal - Satires - 16	267	0.95917
Persius - Satires - Prologus	46	0.93076
Persius - Satires - 1	619	0.95779
Persius - Satires - 2	369	0.89673
Persius - Satires - 3	573	0.95954
Persius - Satires - 4	234	0.94329
Persius - Satires - 5	923	0.95462
Persius - Satires - 6	381	0.96914
Horace - Satires - 1.1	631	0.98686
Horace - Satires - 1.2	694	0.98289
Horace - Satires - 1.3	706	0.99564
Horace - Satires - 1.4	735	0.99447
Horace - Satires - 1.5	499	0.97280
Horace - Satires - 1.6	679	0.95426
Horace - Satires - 1.7	164	0.97914
Horace - Satires - 1.8	239	0.95385

Horace - Satires - 1.9	414	0.96994
Horace - Satires - 1.10	477	0.99319
Horace - Satires - 2.1	421	0.98489
Horace - Satires - 2.2	694	0.99641
Horace - Satires - 2.3	1657	0.81183
Horace - Satires - 2.4	448	0.99738
Horace - Satires - 2.5	568	0.98051
Horace - Satires - 2.6	612	0.99153
Horace - Satires - 2.7	592	0.95033
Horace - Satires - 2.8	461	0.98550

Table 3.3 Book 28 against individual poems.

Book 28 793 - 851		
Poems	**Poem Length**	**Coefficient**
Lucilius - Satires	246	1.00000
Juvenal - Satires - 1	770	0.88441
Juvenal - Satires - 2	737	0.90470
Juvenal - Satires - 3	1419	0.83870
Juvenal - Satires - 4	671	0.90851
Juvenal - Satires - 5	752	0.92915
Juvenal - Satires - 6	3085	0.79303
Juvenal - Satires - 7	1096	0.94070
Juvenal - Satires - 8	1169	0.94180
Juvenal - Satires - 9	696	0.95375
Juvenal - Satires - 10	1689	0.78185
Juvenal - Satires - 11	946	0.85796
Juvenal - Satires - 12	571	0.84138
Juvenal - Satires - 13	1162	0.88233
Juvenal - Satires - 14	1524	0.89649
Juvenal - Satires - 15	811	0.80383
Juvenal - Satires - 16	267	0.93358
Persius - Satires - Prologus	46	0.88085
Persius - Satires - 1	619	0.97133
Persius - Satires - 2	369	0.97259

Persius - Satires - 3	573	0.97982
Persius - Satires - 4	234	0.97495
Persius - Satires - 5	923	0.99050
Persius - Satires - 6	381	0.91404
Horace - Satires - 1.1	631	0.98499
Horace - Satires - 1.2	694	0.99476
Horace - Satires - 1.3	706	0.96840
Horace - Satires - 1.4	735	0.97024
Horace - Satires - 1.5	499	0.93670
Horace - Satires - 1.6	679	0.86576
Horace - Satires - 1.7	164	0.94537
Horace - Satires - 1.8	239	0.95915
Horace - Satires - 1.9	414	0.92724
Horace - Satires - 1.10	477	0.97593
Horace - Satires - 2.1	421	0.97624
Horace - Satires - 2.2	694	0.97457
Horace - Satires - 2.3	1657	0.91582
Horace - Satires - 2.4	448	0.95112
Horace - Satires - 2.5	568	0.95054
Horace - Satires - 2.6	612	0.97569
Horace - Satires - 2.7	592	0.91168
Horace - Satires - 2.8	461	0.97539

Table 3.4 Book 29 against individual poems.

Book 29 852 - 973		
Poems	**Poem Length**	**Coefficient**
Lucilius - Satires	494	1.00000
Juvenal - Satires - 1	770	0.73332
Juvenal - Satires - 2	737	0.80205
Juvenal - Satires - 3	1419	0.82596
Juvenal - Satires - 4	671	0.79074
Juvenal - Satires - 5	752	0.94075
Juvenal - Satires - 6	3085	0.66134
Juvenal - Satires - 7	1096	0.83004

Juvenal - Satires - 8	1169	0.91252
Juvenal - Satires - 9	696	0.96459
Juvenal - Satires - 10	1689	0.59943
Juvenal - Satires - 11	946	0.76122
Juvenal - Satires - 12	571	0.70318
Juvenal - Satires - 13	1162	0.75998
Juvenal - Satires - 14	1524	0.79880
Juvenal - Satires - 15	811	0.64174
Juvenal - Satires - 16	267	0.89184
Persius - Satires - Prologus	46	0.84894
Persius - Satires - 1	619	0.97162
Persius - Satires - 2	369	0.88411
Persius - Satires - 3	573	0.96447
Persius - Satires - 4	234	0.95370
Persius - Satires - 5	923	0.96395
Persius - Satires - 6	381	0.99261
Horace - Satires - 1.1	631	0.97651
Horace - Satires - 1.2	694	0.96479
Horace - Satires - 1.3	706	0.99072
Horace - Satires - 1.4	735	0.99529
Horace - Satires - 1.5	499	0.92144
Horace - Satires - 1.6	679	0.97152
Horace - Satires - 1.7	164	0.93652
Horace - Satires - 1.8	239	0.88445
Horace - Satires - 1.9	414	0.99673
Horace - Satires - 1.10	477	0.99116
Horace - Satires - 2.1	421	0.98519
Horace - Satires - 2.2	694	0.98840
Horace - Satires - 2.3	1657	0.85296
Horace - Satires - 2.4	448	0.96993
Horace - Satires - 2.5	568	0.99916
Horace - Satires - 2.6	612	0.99542
Horace - Satires - 2.7	592	0.98609
Horace - Satires - 2.8	461	0.98224

Table 3.5 Book 30 against individual poems.

Book 30 1000 - 1130		
Poems	**Poem Length**	**Coefficient**
Lucilius - Satires	555	1.00000
Juvenal - Satires - 1	770	0.82030
Juvenal - Satires - 2	737	0.86227
Juvenal - Satires - 3	1419	0.79198
Juvenal - Satires - 4	671	0.84219
Juvenal - Satires - 5	752	0.92438
Juvenal - Satires - 6	3085	0.76020
Juvenal - Satires - 7	1096	0.89050
Juvenal - Satires - 8	1169	0.91455
Juvenal - Satires - 9	696	0.95582
Juvenal - Satires - 10	1689	0.69205
Juvenal - Satires - 11	946	0.79520
Juvenal - Satires - 12	571	0.76682
Juvenal - Satires - 13	1162	0.81185
Juvenal - Satires - 14	1524	0.83670
Juvenal - Satires - 15	811	0.71967
Juvenal - Satires - 16	267	0.88731
Persius - Satires - Prologus	46	0.82192
Persius - Satires - 1	619	0.98715
Persius - Satires - 2	369	0.96579
Persius - Satires - 3	573	0.98762
Persius - Satires - 4	234	0.98767
Persius - Satires - 5	923	0.99925
Persius - Satires - 6	381	0.93894
Horace - Satires - 1.1	631	0.98029
Horace - Satires - 1.2	694	0.98566
Horace - Satires - 1.3	706	0.96906
Horace - Satires - 1.4	735	0.97615
Horace - Satires - 1.5	499	0.90202
Horace - Satires - 1.6	679	0.88734
Horace - Satires - 1.7	164	0.91562

Horace - Satires - 1.8	239	0.91045
Horace - Satires - 1.9	414	0.95410
Horace - Satires - 1.10	477	0.98058
Horace - Satires - 2.1	421	0.97939
Horace - Satires - 2.2	694	0.97545
Horace - Satires - 2.3	1657	0.93852
Horace - Satires - 2.4	448	0.93790
Horace - Satires - 2.5	568	0.96988
Horace - Satires - 2.6	612	0.98340
Horace - Satires - 2.7	592	0.94674
Horace - Satires - 2.8	461	0.97492

In Table 3.5 Horace 1.4 is always above 0.97000. This poem was quoted by Quintilian above as exemplifying the genre. It was additionally shown to be mentioned frequently in the secondary literature. Since our method has borne out this highly correlated poem, it should follow that other poems that are highly correlated to Lucilius should correspond closely to this genre. We have picked a single poem from these data to do a comparative study. This poem has largely been ignored in reference to a comparative study against Lucilius. It is our firm belief that this poem will yield a profitable comparative study. It is to be stressed that our data is not a silver bullet, but provides clues or hints where to focus our study.

Chapter 4 - Lucilius Book 25 & Juvenal 9, A Comparative Study

Our method has proven valid thus far. We have confirmed through our method what scholars have seen as the highest correlated corpus, namely Horace, and now one of the highest correlated poems, 1.4. Therefore, other poems that are highly correlated in our dataset, but have been ignored by scholars, merit a deeper look. Our method has found a subtle dense intertextuality that has lain latent from scholars because Lucilius exists in fragments. The fragments of Lucilius are mostly one or two lines (94.1%). These lines are often disjointed and confusing to read without the luxury of context. Additionally, Nonius frequently has to explain Lucilius' word choice and odd usages of case. For these reasons, it is not easy to relate two texts when one exists in ostensibly disconnected fragments unless it is done by a computer.

Horace 1.4 has a coefficient of 0.99883 when compared against the 26th book of Lucilius. This is an incredibly high coefficient. We assume that since our method has confirmed Horace 1.4 as one of the highest correlated poems, those poems of our other satirists that correspond with high coefficients should have a subtle intertextuality that has been concealed from scholarship. We turn now to Juvenal 9. When compared against book 26 of Lucilius, it has yielded a coefficient of 0.98007. Our method has focused our attention on not only Juvenal 9, but also the 26th book of Lucilius that is only around 100 lines and 500 words.

Few have compared Juvenal at all to Lucilius, let alone correlated Juvenal 9 against Lucilius' 26th book. Admittedly, Juvenal 9 may have been ignored because of its questionable content, as Highet says it "is one of the most shocking poems ever written (117)." Modern scholarship has sought to rectify this, but few have compared Juvenal to Lucilius, with the exception of Juvenal's opening programmatic poem (J 1.30ff; Anderson, "The Programs of Juvenal's Later Books" 145, 147; Jones 17; Umurhan, *Spatial Representation in Juvenal's Satires* 39; Braund and Raschke 75). Many see Juvenal as distinct from Lucilius in his use of libertas (Anderson, "The Programs of Juvenal's Later Books" 148; Harrington 43; Gellar-Goad 46). Libertas is that distinctive quality of Roman Satire that allows a satirist to attack not only vice, but also the men in question who are prone to that vice. Since Juvenal writes under emperors who could easily banish or kill, he deems it safer to write about the dead (J 1.147-171). By contrast, Lucilius could attack anyone. He enjoyed an almost untrammeled libertas owing to his high social status and the republican climate (Umurhan, *Spatial Representation in Juvenal's Satires* 39). Thus Broder says that there is no direct comparison between Lucilius and Juvenal (91). Highet states Juvenal's content is far removed from all other satirists and is therefore distinct (117). Harrington sees all successors of Lucilius to be distinct from the genre he made popular because, while Lucilius' satires are interpreted directly, his successors are often made to say the opposite of what their texts state (25,26).

More often than not, authors compare Juvenal to those outside the normal satirical canon. Throughout the last hundred years most have seen Juvenal to be similar to Martial (Taylor 362-364; Umurhan, "Poetic Projection in Juvenal's

Satires" 234,235; Colton 49; Hutchinson 32, 139; Williams 28,29,346,347; Boswell 75) with his often ribald manner. Some have noticed a similarity to Petronius' *Satyricon* (Taylor 366; Woods 12; Williams 190), while others have preferred to understand Juvenal's satire more in keeping with Horace's version of satire (Keane "Satiric Memories" 227,228; Jenkyns 35; Ulden 112,113; Anderson 155; Highet 295). Both Highet and Gellar-Goad quote from Lavagnini's work *Motivi diatribici in Lucrezio e Giovenale* and see a correspondence of Juvenal to Lucretius' *De Rerum Natura* (Highet 295; Gellar-Goad 45, 46). There are various other comparisons to be noted: Lucian (Jope 59); Bellandi sees a similarity to Dido's words in Vergil's *Aeneid* (Plaza 494, 495); Pomponius (Williams 83); Ovid (Ulden 103); Roman and Greek Comedy (Highet 118, 119).

Surprisingly, a few authors have indeed compared Juvenal against Lucilius in more than just his programmatic poem. While these instances are few they are nonetheless significant in view of our study. In Braund and Raschke's playful discussion of the Juvenal persona, they say he is an agent of destruction just like Dr. Frankenstein. They both play with the dead in their laboratories (Braund and Raschke 71). Dr. Frankenstein uses dead body parts to fashion a new creature while Juvenal in 1.171 says he will use the dead in his *Satires*. Braund and Raschke further note the Juvenal persona is compromised by his characters of lower moral character and contributes to the moral degradation, as do the readers (67-70). In their discussion of this word monstrum that denotes Dr. Frankenstein's monster and the S*atire* of Juvenal, they reference Juvenal's use of this word in *Satire* 2 and *Satire* 8 and take note of Lucilius' similar language (Marx 1342; Marx 117-118; Braund and Raschke 81). Additionally, Williams sees

an indirect reference to Lucilius in Juvenal 9.133 on the effeminate man who is prone to scratch his head with one finger. Such a man would use only one finger so that he would not mess up his elaborate hairdo (Rudd and Barr 202). In a footnote Williams cross-references Juvenal 9.133 with a Lucilian fragment of a homosexual scene where a lover scratches the head of another gently (Warmington 293; Williams 357). One final reference will be shown below. Miller observes that Juvenal paraphrases a Lucilian fragment (*Latin Verse Satire* 298; see page 94).

While Juvenal and Lucilius share a common genre, they are unlikely candidates for a high correlation because of the social changes noted above. We do however find a marked similarity in not only language, but also in thematic structure. Both authors employ a similar dialogic and didactic structure in the first and second person. This dialogic structure is, at times, a heated exchange between the author and a person who needs instruction or lacks moral character. In addition, both authors throw a sustained negative light upon the institution of marriage. Both authors have many complex allusions to other classical authors. Juvenal even parodies a line of Homer's dactylic hexameter in Greek, thus imitating Lucilius more than Horace with a fusion of Greek and Latin. A crudeness not atypical of Roman Satire, centered on sickness, sexuality and excess exists in these lines of poetry. Last, commerce plays an important role in both author's psychological underpinnings. It is this theme of commerce with which both authors struggle and that is a driving force in their search for morality.

Our method has found that book 26 of Lucilius and Juvenal's ninth satire have similar word vectors. This means that in each of these vectors the words

are so closely matched, it is as if both authors drew from the same lexical palette. While language choice does not always indicate a definite correspondence in theme, our attention is drawn to these lines of poetry and we see a close thematic structure.

A dialogue and a didactic structure exists throughout Juvenal's ninth satire as well as Lucilius' 26th book. This structure is marked by a large number of all personal pronouns (Lucilius (29): 633, 637, 640, 642, 643, 647, 651, 655, 656, 657, 666, 669, 670, 672, 674, 674, 690, 691, 696, 701, 701, 702, 702, 703, 704, 707, 712, 713, 717; Juvenal (34): 1, 3, 14, 32, 34, 45, 49, 55, 70, 75, 76, 80, 82, 86, 90, 91, 92, 109, 112, 121, 129, 130, 132, 134, 138, 140, 142, 143). These personal pronouns in Lucilius become increasingly significant when you compare the total occurrences in all his fragments--only 209. This means 14% of all the personal pronouns in Lucilius' fragments occur in book 26, while it only has 6.5% of the total words (the words in question here are only those in our document vectors; this will not match the absolute total number of words, but only those words that we used based upon lemma data). Juvenal's ninth satire has 34 occurrences of personal pronouns, that is 10% of his total, but it is among those satires containing the most personal pronouns. It is tied with satire 14 with 34 occurrences among 1,524 words in the document vector. Only satire 3 and satire 6 in Juvenal, Persius 5, Horace 1.6 and Horace 2.3 have more instances (a table with these frequencies can be found in Appendix I). Juvenal 9, therefore, is the 6th highest poem in personal pronouns among Juvenal, Persius and Horace. When you compare the total number of personal pronouns to the number of words in each poem's document vector, it is the third highest. (In addition, notice

below, Horace 1.6 has an astounding amount of personal pronouns based upon its document vector--7% of its total words. This fact alone has the makings of another comparative study.)

Table 4.1 Personal pronouns in Roman Satire.

Poem	Document Vector Words	Personal Pronouns	Percentage
Horace 1.6	676	47	7.0%
Lucilius Book 26	498	29	5.8%
Juvenal 9	696	34	4.8%
Horace 2.3	1657	71	4.3%
Persius 5	923	35	3.8%
Juvenal 3	1419	35	2.5%
Juvenal 14	1524	34	2.2%
Juvenal 6	3085	41	1.3%

The data in this chart speaks for itself. There is a high correspondence of personal pronouns in both Lucilius 26 and Juvenal 9. Moreover, there is an incredible likeness between these two authors in their use of these pronouns as will be shown below.

When we compare the actual lines of poetry, it becomes increasingly significant how similar each author uses their pronouns. Our correlation has indicated a significance to the personal pronouns. We now expand our view to include those verbs in the first and second person that are not indicated in the chart above.

When we examine the first person singular pronoun, ego and its corresponding verbs, we see 17 occurrences of ego (and its declension) in book 26 of Lucilius, and 12 occurrences in Juvenal 9. In addition, when we examine

first person singular verbs, we find that they are often used in both authors to express what the subject would wish, or to express deliberations of counsel of what he wants his addressee to do.

At the outset, Lucilius declares he does not wish to be read by the most learned, nor the most unlearned (631-635). Cicero describes Lucilius' reasons for his desires in *De Oratore* II.25, "For just as Lucilius, a learned and extremely urbane man used to say, he desired to write to those who were neither the most inept nor the most learned, because the one group might understand nothing and the other perhaps more than himself." Cicero comments, "for I prefer my speech to be misunderstood than for them to find fault with it" (Translation mine, Warmington 202; Cicero 214,216; Cichorius 104). Much to our chagrin, Cicero does not quote Lucilius verbatim. Nevertheless he gives us great insight into these incomplete fragments. Cicero understood that Lucilius situated himself somewhere along the mean between the most learned and the unlearned. Within 4 short lines, Lucilius uses nolo (I do not desire), volo (I desire) twice and non curo (I do not care for). In these instances he is actually addressing himself to the second person singular reader, i.e. you.

In what Warmington delineates as *Satire* 1 in book 26 (632-646), Lucilius describes a impure household, one of promiscuity, "infidam familiam..inpuram domum" (639). He continues:

> Ferri tantum si roget me non dem quantum auri petit, / si secubitet sic quoque a me quae roget non impetret. / Homines ipsi hanc sibi molestiam ultro atque aerumnam offerunt; / ducunt uxores, producunt quibus haec faciat liberos. / qua propter deliro et cupidi

officium fungor liberum.

If she would ask me, I would not give her as much iron [in place of] how much gold she seeks, / even if she would lay down by herself, she still would not obtain what she asks from me. / Men bestow this trouble on themselves and hardship voluntarily; / they take wives, they bring them forth for whom she makes children. / wherefore I leave the straight and narrow, and perform the free office of desire (642-646, Translation mine).

While this passage does not overtly express a second person subject, it does use the first person singular in a didactic structure. Lucilius says he would not give money to a woman who desires to take his money in the form of plates, goblets, clothing or mirrors (640, 641) in order to spend it on drinking. Perhaps this is the woman of the impure household above. The first person singular subjunctive mood dem in verse 642 indicates Lucilius' wishes. He is speaking about a moral path that should be followed. This is why in verse 646 he says that to lead (take) a wife in marriage is to leave that path. Nonius gives us a translation of how we are to take delirare, "est de recto decedere" (to leave the correct course; Warmington 206). In other words, it is to deviate from Lucilius' assumed course. He deliberates, "where is the source of his motivation that makes him leave this path?" The author offers his personal struggle that he sees also in society. Lucilius writes, we men must be crazy (delirare) since we do what is against our own desires.

Warmington divides Satire II at 647-664. Whether this division is to be

respected (Marx divides Book 26 into three satires, Warmington 200), an unmistakable theme exists in these lines around Lucilius' professional life. He begins by musing in line 647, "I may not indeed be convinced that I may abandon my own fields." Whether we abide by Warmington's objections that mutare should indicate a trade, taking this cue from the following lines (650-1, 207) or whether we believe Lucilius means to reliquish his property over to the state, it is clear that Lucilius considered the matter and indicates his conclusion. In the next fragment he tells his hearer that the hearer should be smarter. He must make sure he gets something as he hands over his money. Lucilius again uses nolo to express what he does not want:

> Publicanus vero ut Asiae fiam, ut scripturarius / pro Lucilio, id ego nolo et uno hoc non muto omnia. / At libertinus tricorius Syrus ipse ac mastigias / quicum versipellis fio et quicum conmuto omnia. Indeed, that I would become a tax-collector of Asia or a clerk / instead of Lucilius; I do not desire this. And I would not exchange all things for this one. / But he is a freedman, a Gaul tribe member, a Syrian himself, one who deserves a beating, / with whom I become a shape-shifter and with whom I exchange all things (650-653).

He reasons that he would not desire to be anything other than what he is, namely Lucilius. He would not want to become a tax-collector in spite of its lucrative wages, perhaps because as Cichorius indicates that the risk involved in this business was too great (101-104; Lines 655, 656). Additionally, his unwillingness

to participate in this business may be as Cichorius says: this group mentioned here is the second of the two groups mentioned above--the most learned and the most unlearned (103). He believes these tax-collectors are on the other end of the spectrum--the most unlearned; and therefore he does not desire to be any sort of tax-collector.

One last struggle and desire that Lucilius shares in the first person singular centers around his counsel of a fellow poet or writer. Warmington divides up Satire 5 at lines 689-719 and follows suit with Cichorius that it is addressed to a historian (Warmington 220, 220). Cichorius indicates that Marx's astute observations tell us that this historian is a younger man who is a protegé to Lucilius (109, Line 689). Cichorius disagrees however that all of these verses are addressed to this same man. He sees Lucilius talking as an instructive friend at times to this younger poet, and at other times, because of his sharp tone, dealing with an opponent "Gegner auseinanderzusetzen scheint" (Cichorius 109, 110). Whether we have two addressees or one, it is clear that Lucilius has a struggle in his heart and wishes to express this to his opponent or protegé.

> Tuam probatam mi et spectatam maxume adulescentiam. / Haec tu si voles per auris pectus inrigarier. / Ego si, qui sum et quo folliculo nunc sum indutus, non queo... / Homini amico et familiari non est mentiri meum. / Mihi necesse est eloqui, nam scio Amyclas tacendo periise. / Metuam ut memoriam retineas.../ Evadat salem aliquid aliqua quod conatus sum. / Veterem historiam, inductus studio, scribis ad amores tuos. / et quod tibi magno opere cordi est, mihi vehementer displicet. / Ut ego effugiam quod te in primis

cupere apisci intellego. / Summis nitere opibus, at ego contra ut dissimilis siem.

And having examined your youth and thoroughly considered it / if you will desire these things to water your breast [from your tears] through hearing. / If I, who I am, and in which sack now I am clothed, I am not able to... / It is not for me to lie to a friend and a familiar man. / It is necessary for me to speak, for I know Amyclas to have perished from keeping silent. / I fear lest you retain the memory... / May something come out from something because I tried / Being led in by eagerness, you are writing an ancient history to your lovers / and because it is in your heart to do this great work, / it is exceedingly displeasing to me / just as I shall flee from what I understand you to especially desire to obtain / you press on to this highest work, but I [am] against this, just as I am different (689-691, 695-703).

It seems reasonable to assume that Lucilius' use of the second person singular pronoun indicates he is talking to someone specific. He has considered what he is about to say. He has deliberated thoroughly in order to counsel this young man on what style of writing he is to pursue. He poetically asks this protegé to consider what is being said, that he would let his heart be malleable in Lucilius' hand, "if you are willing by these things to irrigate your chest [with your tears] through [what is said] in your ears (690)." Lucilius describes himself as being clothed in his poetry, being inextricably linked to it. In fact, for him to keep silent would mean peril for his soul. This is incredibly displeasing to Lucilius and he

must counsel him against it. He further instructs his addressee personally on what to avoid (712), what to reason (707), to what object he should devote himself (717) and what to esteem (718).

Similarly, within the complex frame of Juvenal 9 the first person alternates between Juvenal himself and the immoral client Naevolus. While this didactic structure is similar, it is also perplexing because we do not know exactly who is instructing whom. What is clear is that instruction is happening. In the first two lines this is clear, "Scire velim quare totiens mihi, Naevole, tristis / occurras fronte obducta ceu Marsya victas, I desire to know why so many times, Naevolus, / you meet me being sad with a cloudy face just as the defeated Marsyas (1,2)." It is not unreasonable to assume this request for information inherently indicates this behavior is inappropriate. He counsels his subject to not only give an answer, but also to reform himself. The second person pronouns/verbs endure in lines 3, 9, 12, 13 and 14 where he explains the grim face is now what was once the opposite. Naevolus used to be content with a little bit, "agebas contentus modico (9)," but now he is the exact opposite in all ways "omnia nunc contra (12)."

At line 27 Naevolus takes over the first person and replies to Juvenal. Naevolus is distraught because he gets no reward for his services rendered, "at mihi nullum inde operae pretium (27,28). Naevolus blames his bad fortune on the fates themselves, "fata regunt homines" (32). He states a truism that he expects Juvenal to accept. He instructs Juvenal that if the fates have determined you to fail because they have left your side, then no resources in your arsenal can help you "nam si tibi sidera cessant, nil faciet longi mensura incognita nervi (33,34)." The fate allotted to Naevolus is revealed--his patron is cheap (38ff).

Juvenal continues at line 47 by telling Naevolus to remember his past. He uses the familiar language we have seen throughout Juvenal and Lucilius of deliberating. Both use expressions of cognition and discernment through a variety of words: scio, nescio, puto, intellego, cognosco, dubito, suadeo, etc. Here Juvenal says, "Sed tu sane tenerum et puerum te et pulchrum et dignum cyatho caeloque putabas" (but you however used to consider yourself a tender and pretty boy and worthy of being a cupbearer in heaven J47, J48)." He instructs him by offering Naevolus' own words that he thought at one time, perhaps these words will make him change his current view.

Naevolus' response to Juvenal's prompting for reasoning is startling. Naevolus uses the second person plural pronoun indicating that Juvenal's persona is in the same category as his cheap patron.

> Vos humili adseculae, vos indulgebitis umquam / cultori? Iam nec morbo donare paratis? / En cui tu viridem umbellam...mittas
>
> Will you (plural) ever be kind to your humble follower? Will you ever be kind to the one who ploughs you? / Do you now not prepare to bestow something for your disgusting gratification? / But to whom you might send a green umbrella (48-50).

Naevolus begins by saying his patron and Juvenal are the same kind of patrons, those who never treat their clients well. They are both cheap, unwilling to bestow gifts even for the curing of their disease. Naevolus then turns to address his patron specifically as he uses the second person singular pronoun. In fact, we do not even see another second person plural verb until line 69, "Durate atque

expetate cicadas" that is addressed not to his patron and Juvenal, but his own slave boys. One cannot escape such a pointed rebuke in the mouth of Naevolus toward Juvenal. Naevolus, who is supposed to be the one deficient in moral character, rebukes his patron and Juvenal for a lack thereof. Perhaps this harkens back to Lucilius' encouragement that he and his audience should "munifici comesque amicis nostris videamur viri" (let us seem to our men and our companions bountiful; 657). It is to be noted that Lucilius' verse is among peers while Naevolus is clearly talking about the patron/client relationship.

Naevolus begs Juvenal to keep silent. Second person pronouns are used throughout his plea in lines 92-101. Naevolus fears for his life if his secrets are told. Ironically, if this poetry in Satire 9 were real events, Juvenal has not kept silent since we are now reading it (Hutchinson 138). Juvenal counters that no rich man can ever keep a secret because they live such a public life, their doings will always be known to everyone (107ff). Even if the master is innocent their servants will concoct stories (110, 111). In the end, Juvenal counsels Naevolus to live a proper life so he can ignore the tongues of his slaves and never worry about his secrets, since there would be no secrets to be told (118). Naevolus himself has also shared his secrets and as Braund astutely observes he is no better than the slaves he repudiates for telling secrets (Miller, *Satiric Grotesques in Public and Private* 67,68).

Naevolus admits Juvenal to be counseling him as he says "utile consilium modo, sed commune, dedisti," (you have given me useful counsel just now, but it is general). He wants further advice from Juvenal to tell him exactly what to do right now, "nunc mihi quid suades (125)." Juvenal moves away from his advice

above to live a proper life (110,111) and ostensibly jeers at Naevolus by saying he will never be without a pathic patron. Juvenal ends his poetic admonishment by describing an effeminate army coming from all corners of the globe, in both carriages and ships. This army, instead of having a motive to conquer, will come in order to submit themselves passively to Naevolus (131-133). It may be that Juvenal is turning the language of battle, seen so much in book 26 of Lucilius, on its head (708, 709, 710, 714, 715, 731, 732, 734).

Both authors quite significantly throw a sustained and noticeable negative light upon marriage. Marriage is either an obstacle to be avoided in Lucilius, or a simply a law to be circumvented in Juvenal. Our method has pinpointed our focus to both of these areas because of similar familial words such as coniunx (Lucilius 639, Juvenal 79), mater (Lucilius 704; Juvenal 23, 60), uxor (Lucilius 645; Juvenal 71) and domus and domina (Lucilius 639; Juvenal 79). Because our method has attracted our gaze to these passages, we now see strikingly similar language in meaning with the phrase producing children (Lucilius: faciant liberos, 645; Juvenal: filiolus...filia nascitur, 82) and an unfaithful relationship in a marriage or household (coniugem infidamque pathicam familiam inpuram domum Lucilius, 639; coniugium in multis domibus servavit adulter, Juvenal 80). These examples are significant parallels whether consciously done or not. In fact, when you examine Lucilius' other fragments there is no other such sustained passage on marriage with this vocabulary. Let us examine Lucilius and Juvenal's attitude toward marriage.

Lucilius begins the context of the passage quoted above (642-646) by talking about a trip he had recently taken on foot (repedabam). We are left to

speculate on the exact context, but perhaps in coming back from his destination on foot he stopped at an inn where he might have noticed these normal household implements, cribrum, lucerna, tela, later and licium (a sieve, a lantern, a warp of a loom, a brick and a thread). In another context Lucilius uses the word lucerna along with the word bed (lectus), that is surely one of the most common household items (16). While we are at a loss to know the exact context of these implements, it seems fairly reasonable to assume he is describing a normal Roman domus (household). The very next fragment we are given by Nonius is line 639 where Lucilius describes "a spouse, an unfaithful promiscious household, an impure home." This normal Roman household had become polluted by unfaithfulness, unfaithfulness that in some way threatened to involve Lucilius himself.

> depoclassere aliqua sperans me ac deargentassere / decalauticare eburno speculo despeculassere. / Ferri tantum si roget me non dem quantum auri petit. / si secubitet sic quoque a me quae roget non impetret
>
> Some lady hoping to de-goblet me or de-silver me / or to deprive me of women's shawls or de-ivory mirror me / If she would ask me, I would not give her as much iron [in place of] how much gold she seeks, / even if she would lay down by herself, she still would not obtain what she asks from me. (640-643).

Could it be that some innkeeper's wife had propositioned Lucilius, or that Lucilius means to suggest this in his poem? This could have happened in his stay at the

inn while her master was away. Lucilius could already tell this household was infected by some impurity as Nonius indicates (Warmington 204). The spouse further fouls her reputation by trying to strike a deal with Lucilius. He is poetically descriptive of her intense avarice, she wishes to take all Lucilius has (641). He, however, seems to be unmoved in her request. He would not even give her iron for as much gold as she had asked because her offer is so odious to him. Lucilius further exclaims this offer is so repugnant, and she is so polluted that he desires to have no further dealings with this adulteress. He would not even strike a deal with her, that she might go away and sleep by herself, for even this would besmirch his character (Pereira reconstructs this passage by stating the one lying down is the husband trying to avenge himself from his adultress wife, i.e. depriving his wife 23, 24). It is in this context, Lucilius states marriage is a nuissance or an annoyance to be avoided. For it is men themselves who have created their own burdens (aerumna) in their taking of wives (ducunt uxores, 644). Men have gone insane, or have stopped plowing straight when they take wives in their desire for children (645; Warmington 206, 207). Marriage is treated as a vice by Lucilius to be avoided for it is seemingly the root cause of men's problems.

In like manner, Juvenal shows in *Satire* 9 that Naevolus believes marriage to be not only a vice, but only a charade in order to glean benefits from the state by having children.

> ...uxor tua virgo maneret? / Scis certe quibus ista modis, quam saepe rogaris / et quae pollicitus. Fugientem nempe puellam / amplexu rapui; tabulas quoque ruperat et iam / signabat; tota vix

> hoc ego nocte redemi / te plorante foris. Testis mihi lectulus et tu, / ad quem pervenit lecti sonus et dominae vox. / Instabile ac dirimi coeptum et iam paene solutum / coniugium in multis domibus servavit adulter / ...quod tibi filiolus, quod filia nascitur ex me? ...your wife would still be a virgin? / Truly you know in what way, and for what you asked so often and what was promised. / Certainly your girl was fleeing / when I snatched her in an embrace; she had also destroyed the tablets and now / was making a new signature of marriage; I recovered [your marriage] through the night / while you were crying at the doors. The little bed and you were my witnesses, / to whom the sound of the bed and the sound of your mistress came straight away. / In many households an adulterer saved the day, / with a nearly dissolved union, an unstable marriage and one that has started to break up / because your little son or your daughter is born from me? (J71-79; 82)

Ironically, Naevolus paints himself as the only faithful and devoted (devotus... deditusque 71, 72) member of the household/marriage of his patron. While this wife married him in good faith, Naevolus' patron was unwilling or unable to consecrate his marriage. In Satire 9, his affections appear to be otherwise occupied as the passive member with Naevolus (27-46). In quite graphic terms we are made aware that Naevolus' patron is quite disinterested in his wife and he desires to be dominated sexually by his client (43,44). Many marriages, we are told by Naevolus, would end in dissolution if not for an adulterer to impregnate

the mistress. The adulterer is actually the person who saves marriages in Naevolus' view. We cannot say that Naevolus views all marriages in this way since he qualifies his statement, "in multis domibus," but we can however say Naevolus' view of marriage from his own experience with his patron is quite dismal.

The descriptions of marriage in Lucilius and Juvenal are similar. Could it be that Juvenal uses Lucilius' *Satire* in book 26 for his characterization of Naevolus? The greed portrayed by Lucilius' woman of low morals parallels that of Naevolus. Lucilius describes the unfaithful spouse as one who desired not only his silver, but his cups, his shawls and even his ivory mirrors. Lucilius would not give her as much iron as she asks for in gold, perhaps, not because he is so unwilling to pay, but perhaps because she asks for such an extravagant amount. Juvenal is no less avaricious. "Naevolus' list of necessities is extravagant. He expects an income (fenus) just below the equestrian census from his property, a silver plate, litterbearers to take him to the circus, an engraver and a painter. This is far more than is necessary to meet the needs of the venter" (Miller, *Latin Verse Satire* 304, 305). Naevolus, while basically a slave (Miller, *Latin Verse Satire* 305; Juvenal 45) wishes to become part of the aristocratic elite in Rome simply from his gigolo practices (Miller, *Latin Verse Satire* 305). The greed that Naevolus is expressing through the pen of Juvenal is beyond excessive. He expects to switch places with his patron. Thus, both the innkeeper and Naevolus' patron show an excessive greed, both in the context of payment for their unfaithfulness.

It is significant that both authors have similar allusions to other classical authors. Lucilius quotes from Pacuvius many times, the famed tragic poet of

Rome. Juvenal follows Lucilius' paradigm closely and makes many references to other literary works, e.g. Vergil. This shows both authors weave complex allusions throughout their poems. Both employ three clear allusions to Homer. Juvenal even offers a parody of a line of Homer's *Odyssey*, thus emulating the hybrid of Lucilius even more closely than Horace since Horace stated that to mix Latin and Greek is not a high achievement (Horace 1.10.25ff).

Lucilius is replete with not only the Greek tongue, but also references to Greek authors. Allusion for Juvenal is important as well. It is quite significant that in such a short span of 100 lines, both authors use Homer's epic poetry in their satire. Warmington delineates 10 lines within book 26 as *Satire III* (665-675). Line 665 begins with an allusion to Agamemnon that is not an overt allusion to Homer, but suggests a Homeric reference nonetheless. While Warmington quotes Fiske who believes it is a clear reference to Pacuvius, it is understood that ultimately this name can be found in Homer. In other words, the mere mention of this name in any form is a reference to Homer. The words "Ego enim contemnificus fieri et fastidire Agamemnonis," "for I become scornful and disdain Agamemnon" sound like words that Achilles would utter (666; Warmington 215). Immediately after this is a reference to Athena's anger against Ajax, "nec Minervae prosperatur pax quod Cassandram...signo deripuit," "nor is the peace of Minerva rendered favorable because he ripped Cassandra away from the statue" (667, 668). Lucilius has Homer in view as Ajax would have escaped death even with all of Athena's wrath against him, if not for his boastful words that the gods could not drown him (*Odyssey* 4.500ff). Lucilius' second allusion to Homer is at line 733, "Solus illam vim de classe prohibuit Vulcaniam,"

(he alone held back that Vulcan force from the naval fleet)." The reference is again to Ajax standing courageously to meet Hector in battle and the other Trojans who have in mind to burn all the ships of the Achaeans (*Iliad* 15.670ff). While Ajax is disarmed and flees from Hector, only one ship is burned--the fleet is saved. Ajax alone held back this Vulcan force that would have conquered the Achaeans. Another allusion to the *Iliad* is seen in line 734, "Domutionis cupidi imperium regis paene inminuimus," in our desire for home-going we nearly impaired the authority of the king. Warmington sees this as a clear reference to Odysseus' striking of Thersites for his insolence against Agamemnon (*Iliad* 2.210ff). Thersites counsels all the Greeks to leave Agamemnon alone and sail for home. Odysseus castigates Thersites for his impudence against the king, warns him that he will utterly humiliate him if he does this again and then hits him on the back and shoulders as a warning (265ff).

Of the six overt allusions in Juvenal three are from the *Odyssey*. One allusion is simply a reference to a name, while another is a complex allusion where Juvenal even parodies a Homeric line in Greek. One cannot help but think Juvenal was closely following the allusions in book 26 of Lucilius. While it could be a coincidence that in the sparse 150 lines of *Satire* 9, Juvenal randomly quotes from Homer, it is interesting nonetheless.

The first allusion is on line 37 and it is one of the most complex allusions to Homer in all of Roman satire. Juvenal parodies a specific line of Homer.

> ...et blanae adsidue densaeque tabellae / sollicitent,
>
> αὐτος γάρ ἐφέλκεται ἄνδρα κίναιδος.
>
> his flattering and frequent letters constantly / stir you

up, for the catamite himself attracts man (36,37).

The original line in the context of Homer's epic is thus:

> ἐκ καπνοῦ κατέθηκ᾽, ἐπεὶ οὐκέτι τοῖσιν ἐώκει / οἷά ποτε Τροίηνδε κιὼν κατέλειπεν Ὀδυσσεύς, / ἀλλά κατήκισται, ὅσσον πυρός ἵκετ' ἀϋτμή. / πρός δ' ἔτι καί τόδε μεῖζον ἐνί φρεσί θῆκε Κρονίων,
> μή πως οἰνοθέντες, ἔριν στήσαντες ἐν ὑμῖν, / ἀλλήλους τρώσητε καταισχύνητέ τε δαῖτα / καί μνηστύν; αὐτός γάρ ἐφέλκεται ἄνδρα σίδηρος.'
>
> "I placed [it] away from the smoke, since it is no longer what it used to be, / such as when Odysseus went away, going away to Troy. / But it is spoiled, as much as it has come to the fiery breath of the bellows," / but to them yet also then say, "the son of Kronos may place it in their minds, / lest being in this way intoxicated, would stir up strife among you, / that you may wound one another, and dishonor the marriage feast, / for an iron weapon itself attracts man (*Odyssey* 16.288-294)."

Odysseus directs his son, Telemachus in this passage to gather up all the armor and put it into the store room. He directs him to lie to the suitors if anyone asks, by saying he wishes these armaments to be out of site because in their drunken state they would be even more susceptible to the general principle that "arms themselves attract men [to use them]." Miller notices the 'deep attraction of violence" in such a statement (*Latin Verse Satire* 299). The shear vision of arms incites men to violence. This principle is changed thus, "a catamite incites men

to dominate him." Miller sees the missing Greek σίδηρος (iron, arms) was added as the Latin sidera (stars) in line 33 as a near-homophone to indicate that in spite of the stars abandoning Naevolus, he will control his master's fate through his dominance. The sound of these words are so similar that one cannot dismiss this suggestion even if one disagrees with Miller's interpretation. Or perhaps it was Juvenal's unwitting word choice of sidera in line 33 above that, when it was composed by Juvenal, put him in mind of the Homeric line. What is perplexing about this quote is that it shows Naevolus as the one out of control, or being enticed by the pathic patron or the catamite. This is reversing the already reversed patron/client relation since Naevolus is pictured elsewhere as the one who has captivated his patron and renders his dominating service to him, and looks for others as well who desire this type of relation (28; 36; 42; 45; 70ff; 92,93; 130-134). In fact, if Naevolus is the one who is captivated, why is there a commerce exchange at all? And in fact, since Naevolus has indicated he has not been paid, there is no commerce exchange. It seems as though the relationship of patron and client in Juvenal 9 is more complex than we think as Miller indicates. It was not strictly a relation of commerce, but a "spontaneous friendship founded on mutual good offices (Miller, *Latin Verse Satire* 301)." The client, Naevolus, dominates the patron and turns the normal Roman social relationship on its head.

The second allusion is in line 64 and 65.

> "improbus es cum poscis' ait. Sed pensio clamat, / 'posce.' Sed appellat puer unicus ut Polyphemi / lata acies per quam sollers evasit Ulixes.

'You are wicked when you ask,' he says, but my rent shouts, / 'Ask!' And my only slave boy calls out just as big-eyed Polyphemus' / does through which the clever Ulysses escaped.

This reference at first glance seems wholly unrelated to our text until you read the entire context of the passage in Homer. The description of Odysseus heating up the fiery point of his stake (πυριήκεα μοχλός) and plunging it into Polyphemus' eye is quite descriptive. It pictures a blacksmith plunging an axe into water after newly forging it.

ὡς δ᾽ ὅτ᾽ ἀνὴρ χαλκεὺς πέλεκυν μέγαν ἠὲ σκέπαρνον / εἰν ὕδατι ψυχρῷ βάπτῃ μεγάλα ἰάχοντα / φαρμάσσων: τὸ γὰρ αὖτε σιδήρου γε κράτος ἐστίν / ὣς τοῦ σίζ᾽ ὀφθαλμὸς ἐλαϊνέῳ περὶ μοχλῷ. / σμερδαλέον δὲ μέγ᾽ ὤμωξεν, περὶ δ᾽ ἴαχε πέτρη, / ἡμεῖς δὲ δείσαντες ἀπεσσύμεθ᾽:

But just as a smith dips an axe or a great blade into cold water / tempers it with a great cry: / indeed this iron is strengthened just as the eye hissed around the olive stake. / He wailed frightfully and he shrieked around the rock-cave, / and we being fearful ran off (*Odyssey* 9.391-396).

The context of the Homeric story is used to emphasize Naevolus' rent crying out, and later his single slave boy calling out. These both cry out exactly like Polyphemus' eye hisses and sputters. Notice additionally in line 393 we see the exact word that is replaced with κίναιδος in line 37, σιδήρος. It is clear that since Naevolus' rent cries out, and the only means of escape--as Odysseus had only

one means of escape--is to statisfy his patron, similarly to the blacksmith who dips his smoking tool into cold water to harden it. This reference, therefore, is quite complex. Juvenal's use of it shows not only his profound knowledge of Homer, but his subtle use of allusion to bring additional meaning to his satire. Even mentioning Odysseus' name might put his readers in mind of Eurylokhos' characterization of Odysseus as a man of iron (*Odyssey* 12.280, ἦ ῥά νυ σοί γε σιδήρεα πάντα τέτυκται) that harkens back to this deleted word in line 37!

Juvenal's last reference to Homer comes at the close of his *Satire*. Naevolus shows himself to be quite ridiculous with his outrageous request of becoming one of the wealthiest Roman citizens by simply pleasuring a patron. He ends by saying that whenever he prays to the gods or the fates, they plug their ears just like Odysseus' crew in book 12 of the *Odyssey* to avoid the Sirens. In this reference, Naevolus has become the Sirens whose alluring petitions mean destruction for the hearers. It is also significant that these warnings come not from Naevolus' interlocutor, but from Naevolus himself. Juvenal puts into the mouth of Naevolus his own destructive tendencies. For Odysseus' men when they saw him signaling to unloose the ropes, rowed faster and tied him even tighter, recognizing the great threat (*Odyssey* 12.192ff). Or is Naevolus' point simply that the Fates ignore him as a deaf man cannot hear? As can be seen by these allusions, both Lucilius and Juvenal use Homeric references weaving the master bard throughout their *Satires*.

There is in both authors, a crudeness that is not atypical of Roman satire. We see this crudeness in the form of sickness, sexuality and excess. We have already brought out the excess found in Juvenal through Naevolus and the

unfaithful spouse found in Lucilius. Their greed goes beyond the ridiculous. Additionally, we have already seen the crudeness in sexuality between these two characters. Moreover, there are many parallel examples of excess mostly centered around drinking and feasting in both Juvenal (10; 113; 116, 117; 128) and Lucilius (654; 658; 659; 664; 665; 722; 727,728; 731). Juvenal even has an example of crudeness that mentions sickness (the bowels), sexuality and excess (feasting) in one thought (42-44). But by far, the most significant parallel examples of crudeness between Lucilius and Juvenal are found in sickness.

Juvenal's theme of sickness is found in multiple passages of his poem (10-21; 42-44;) while Lucilius' is located squarely in mostly lines 678-687. This theme of sickness is found in one preserved passage of Lucilius, but Juvenal uses similar language in his description of Naevolus' sickness from a variety of Lucilian lines (sicco, L688 & J11; aegrotus, L692 & J18; vetus, L700 & J16; squalitas, L729 & J15; dolor, L679 & J89). The language in these passages alone makes us comprehend why our correlation coefficients are so high. In his Roman Satire anthology and reader Miller astutely observes that Juvenal even paraphrases a line from book 26 (*Latin Verse Satire* 298). It is this comment that confirms organically what we have found programmatically.

> Animo qui aegrotat videmus corpore hunc signum dare;
>
> We see he who is sick in his mind gives off this sickness as a sign with his body (L678 Miller is using Krenkel's text).
>
> Deprendas animi tormenta latentis in aegro / corpore, deprendas et gaudia; sumit utrumque / inde habitum facies.
>
> You can discern the torments of the hidden soul in a sick / body,

you can discern also joy; / there the face takes up both conditions (J18-20).

While there are only three words in these lines that indicate a similarity (corpus, animus and aegroto), the meaning of these lines state exactly the same thing. There is absolutely no misunderstanding that Juvenal had not only read Lucilius, but is imitating him in this satire. We have here not only a significant intertextuality between Juvenal and Lucilius, but a paraphrase, as Miller indicates, of Lucilius within the lines of Juvenal. In all of modern scholarship there are few instances where Lucilius and Juvenal are directly correlated, least of all from book 26 of Lucilius to *Satire* 9 of Juvenal. Yet, both my advisor has observed organically what I have observed programmatically. This did not happen by design on my part. I had no preference as to which poem or author I wanted to compare. I will admit, I did have a preference to use Persius or Juvenal, simply because they were more far removed from Lucilius in date, and Horace has always been highly correlated to Lucilius. I only remembered the content of Juvenal 9 after I had decided to use this satire for my intertextual study, and started to read it again.

Finally, commerce is a significant theme throughout both authors. Commerce is a driving force of all players within the satires. It is money that is either embraced as all-powerful and becomes the goal for those prostituting themselves (L639-644; J135-150), or it is alternately forsaken and classified as unimportant for those who pursue morality (L650,651,656,657; J102-123).

As shown in Appendix D, there are many other poems that urge us to

perform an intertextual study based upon the results of our method. Juvenal's Satire 12 against Book 25 of Lucilius has a highly significant correlation coefficient of 0.99133. Persius Satire 6 has a correlation coefficent of 0.99261 against book 29 of Lucilius while Horace 1.4 has a coefficient of 0.99529. Book 30 of Lucilius is highly correlated against Horace 1.4 with a coefficient of 0.97615, while Persius Satire 5 is 0.99925. These correlation coefficients indicate a profitable intertextual study. While looking at all of these poems is beyond the scope of this dissertation, future study could be done on all of those poems that have a coefficient above 0.97.

Chapter 5 - Situating the Dubious Fragments

Since our only knowledge of Lucilius comes from the text of Nonius, and there are variants of his text that exist; we are unsure about lines 974-980 and 981-999. Lines 974-980 could either belong to book 28 or 29 as two of the variants differ. Lines 981-999 could belong to a book within 26-29. We should be able to use our method to see if these lines correlate closely to any one book, hoping that we find a high coefficient in the books we believe the fragments belong. For lines 981-999 our coefficients are as follows (full coefficients can be found in Appendix G):

Table 5.1 Dubious fragment coefficients (lines 981-999).

Books	Poem Length	Coefficient
Lucilius - Book 3	185	0.98543
Lucilius - Book 4	155	0.96236
Lucilius - Book 5	246	0.96885
Lucilius - Book 7	127	0.99205
Lucilius - Book 8	78	0.99273
Lucilius - Book 10	43	0.96702
Lucilius - Book 11	113	0.98965
Lucilius - Book 14	115	0.99106
Lucilius - Book 15	156	0.97994
Lucilius - Book 17	62	0.98852
Lucilius - Book 19	58	0.99008
Lucilius - Book 27	278	0.96000
Lucilius - Book 28	246	0.96795

Only books with coefficients above .96 are displayed above. We do find a relatively high coefficient in books 27 and 28. With this data we may be able to exclude the possibility that this fragment belongs in books 26 or 29. Because we have high coefficients across many books, we may not be as confident as we could be. As for 974-980, the coefficients are below. We expect to see high coefficients in either book 28 or 29.

Table 5.2 Dubious fragment coefficients (lines 974-980).

Books	Poem Length	Coefficient
Lucilius - Book 28	246	0.97236
Lucilius - Book 29	494	0.82506

We do indeed have a high coefficient in book 28 and the difference between this coefficient and book 29 is quite stark. Additionally the language does not glean high coefficients across many other books either. With this data, I believe we can confidently say lines 974-980 belong squarely in Book 28.

Last, we use our method against the unassigned fragments of Lucilius, of which we have no indication or hint as to what book they belong. If we are able to find individual poems that are highly correlated against Lucilius, we should be able to correlate unassigned fragments against the books of Lucilius. Our method will discover to which book a few fragments could be assigned. I expect many of the fragments of Lucilius that we will try to categorize will not give us any clear indication of where they belong, but perhaps a few fragments will yield some interesting values.

We took all of the unassigned fragments of Lucilius (1131-1272) separately to obtain coefficients against the individual books of Lucilius. A few fragments

yielded interesting results (see Appendix F), but we will only look at fragment group 811 (L1196-1208) against Lucilius' book 15 (L507-543). Fragment group 811 is the longest fragment (92 words) in Lucilius as can be seen from the following table.

Table 5.3 Largest fragments in Lucilius.

Fragment Group	Book	Lines	Meter	# of Lines	# of Words
811	?	1196-1208	Hexameter	13	92
246	10	401-410	Hexameter	10	73
122	5	200-207	Hexameter	8	58
117	5	186-193	Hexameter	8	53
774	?	1145-1151	Hexameter	7	52
113	4	176-181	Hexameter	7	48
58	2	87-93	Trochaic septenarius	7	46
345	17	567-573	Hexameter	7	46
523	28	805-811	Iambic senarii	6	41
322	15	524-529	Hexameter	6	41

Because of its length, it will enable us to easily discern if a true correlation exists. At first glace we notice an astounding amount of similar words (see Table 5.4). There are so many words (we are to remember that our stop words are not among these) across relatively few lines of poetry (Fragment 811 is only 13 lines and Book 15 is only 36 lines), that the correlation is quite convincing. It is also significant that within book 15 is fragment group 322. This fragment, that is listed above as the tenth longest fragment within Lucilius, will give us the necessary context to ensure our correlation holds water. It is important to point out that Fragment 811 and book 15 are both written in hexameters. Even though Lucilius

used a variety of meters (sometimes within the same books), it would be a tenuous argument to place an unassigned fragment written in a specific meter into a book which had no fragments with this same meter.

Table 5.4 Similar words in fragment 811 and Book 15.

Word	Verse from Fragment 811	Verse from Book 15
facio	1206	541
homo	1198,1199,1204,1205	519,520,527,535
magnus	1206	513,522,523
malus	1200,1204	523
pretium	1196, 1202	538
primus	1207	519,521,531,538
puto	1207	521,528
scio	1198,1199,1201	542
tertius	1208	539
tertius iam*	1208	539
utilis	1199,1200	508
verus	1196	528,529
verso	1197	513
vivo	1197,1206	527

While no one would deny the similarity in word choice, although some of these words are fairly common, is there a similar theme in both? Can this unassigned fragment be happily situated somewhere in the context of book 15? The beginning of book 15 is devoted to horses (507, 511-513, 514, 515, 516-517, 518) as Warmington states, but he also writes about philosophy in book 15 (162). He then writes about foolish men who believe the superstitions in Homer and declares that all paintings and statues of the gods are simply artists' renderings and they are not real (519-529). Book 15, according to Warmington's

arrangement ends talking about misers (530-543). It is in this context, I believe we can situate fragment 811.

> Praeter quam in pretio; primus semisse, secundus / nummo, tertius iam pluris quam totus medimnus.
>
> On account of the price, first it was sold for a half a bronze pound, second / a sesterce now third more than a Greek bushel (538, 539).

Lucilius describes the price of food (perhaps because of the context of 536,637, Warmington 162) on the cusp of dealing with misers. He uses the ordinals, primus, secundus and tertius. He does this again later in our unassigned fragment though replacing secundus with deinde.

> Virtus, Albine, est, pretium persolvere verum / quis in versamur, quis vivimus rebus, potesse, / virtus est homini scire id quod quaeque habeat res, / virtus, scire, homini rectum, utile quid sit, honestum, / quae bona, quae mala item, quid inutile, turpe, inhonestum, / virtus, quaerendae finem re scire modumque, / virtus, divitiis pretium persolvere posse, / virtus, id dare quod re ipsa debetur honori, / hostem esse atque inimicum hominum morumque malorum, / contra defensorem hominum morumque bonorum, / hos magni facere, his bene velle, his vivere amicum, / comoda praeterea patriai prima putare, / deinde parentem, tertia iam postremaque nostra.
>
> Virtue, Albine, is truly to be able to pay a price in the business / which we move about and live. / Virtue is to know what a matter may hold for a man, Virtue is to know what is upright for a man, /

> what may be useful and honorable, / What is good, likewise what is bad, what is not useful, disgraceful and not honorable, / Virtue is to be striving for the end, and to know the way of a thing, / Virtue, is to be able to pay the price of riches, / Virtue, it is to give as far as itself is owed to the honor of a matter, / It is to be an enemy and unfriendly of evil men and manners, / A defender against good men and good manners, / to hold these things in high esteem, / to be willing for these things well, to live as a friend to these things, / in addition, to think first the advantage to our homeland, / next parents, third now and last our own (L1196-1208).

This resemblance is striking, as is the context of our unassigned fragment describing virtue as the ability to pay the price of riches. At the core of a miser is selfishness, the reverse of this is what Lucilius describes at the end of his description on virtue. To be virtuous is to think first of others, namely our homeland, next our familial relations and finally our own interests. While there is no use of the word virtus in Book 15, and the use of this word would have completely sealed my argument, I believe there is enough context at the very least to say my conjecture is not wholly unreasonable.

Finally, let us use our roving correlation to see where this unassigned fragment correlates highest throughout book 15. Perhaps it will give us an indication or a confirmation where to place fragment 811.

Table 5.5 Fragment ID 811 against Book 15 (Rank-11).

Fragment ID 811 against Book 15 (Rank-11)	
Roving Correlation Lines	**Coefficient**
Lucilius, Satires 313 507-519	-0.02321
Lucilius, Satires 313 508-520	-0.02732
Lucilius, Satires 314 509-521	-0.12070
Lucilius, Satires 314 510-522	-0.12070
Lucilius, Satires 315 511-523	-0.04560
Lucilius, Satires 315 512-524	-0.09522
Lucilius, Satires 315 513-525	-0.08005
Lucilius, Satires 316 514-526	**0.15659**
Lucilius, Satires 317 515-527	-0.21856
Lucilius, Satires 318 516-528	-0.02608
Lucilius, Satires 318 517-529	0.00659
Lucilius, Satires 319 518-530	**0.23231**
Lucilius, Satires 320 519-531	**0.24024**
Lucilius, Satires 321 520-532	-0.01349
Lucilius, Satires 321 521-533	-0.05308
Lucilius, Satires 321 522-534	-0.04453
Lucilius, Satires 321 523-535	-0.10002
Lucilius, Satires 322 524-536	-0.02962
Lucilius, Satires 322 525-537	-0.01570
Lucilius, Satires 322 526-538	-0.02468
Lucilius, Satires 322 527-539	0.01267
Lucilius, Satires 322 528-540	-0.05329
Lucilius, Satires 322 529-541	-0.07605
Lucilius, Satires 323 530-542	-0.08160
Lucilius, Satires 324 531-543	-0.09509

As you can see from our roving correlation, there are two places in book 15 that correlate relatively high with fragment 811. These two places are within 518-531 and with a lesser coefficient lines 514-526. 518-531 is more significantly

correlated than the latter and begins the section on misers. This is precisely the context into that we have situated Fragment 811 above. It is to be noted that this roving correlation is done on fragments and therefore could be much more profitable with a complete text. Finally, it is important to note that Warmington organizes his fragments in an order that is reversed from Marx's edition (Warmington viii, ix). One could argue that we have situated Fragment 811 based upon an erroneous ordering of Book 15. However, the only problem that this argument poses is that our fragment is situated either at the end of the miser context, or it introduces this context. In other words, if the order of book 15 is reversed from Warmington, our fragment concludes what Lucilius has already exemplified with his lines about misers.

The roving correlations in Table 5.5 were done using single lemmata. We have noted correlated words, but they are fairly common words listed in Table 5.4. We have additionally tried correlations using multiple indices (We created document vectors using multiple words instead of single words as described above). We tried indices of 2 through 5 with rank approximations from 1-36 (This data yielded a total of 36 eigenvalues in Σ, therefore we could use a rank anywhere from 1 to 36). Since the data diverged too significantly from the original matrices when it was processed with SVD, it appeared that our coefficients were false-positives. Instead we will process Fragment 811 using a special subject correlation using the words in Table 5.4. A rank-7 approximation was used, but the coefficients changed very little when using other rank approximations thus we are confident in our correlations.

Table 5.6 Lucilius lines 519-530 with differing rank-k approximations.

Lines	Rank-7, I-1	Rank-2, I-2
519-530	0.42645	0.46112

As you can see from Table 5.6 we have a relatively high coefficient in lines 519-530. This coefficient in Table 5.6 are relatively high compared with the zeros and negative coefficients not listed for clarity. These are the same lines above that had a coefficient above 0.24000. When we change our index to 2 words, lines 519-530 has a coefficient of 0.46112. This is remarkably high given that all other coefficients were either 0 or negative.

We have only examined one poem from Appendix D. There were many more than only the four mentioned above on which we could perform a study to determine if the coefficients were false-positives. While it is admitted freely that false positives in our data can occur, and that much more study needs to be done, we have sufficiently shown that our method can determine dense levels of intertextuality between two texts--even with incomplete texts and smaller fragments. This method is independent of language; and therefore, can be used to correlate Classical texts in Greek, as well as any other language.

Chapter 6 - Conclusion

We have shown how important the field of classics has been as a pioneer in the digitizing of documents that had led to more digital document projects. Digital documents are now a multi-billion dollar industry with companies like Amazon and Apple. It has also revolutionized the way research is done today across all disciplines. While other disciplines have been using statistical methods for quite some time, the field of classics and comparative literature has largely ignored using math to test the similarity of documents. In order to bridge this gap we have applied statistical methods to find similarities in classics and comparative literature. We have proposed a method to easily identify similar texts from multiple authors in order to transfix our gaze to the most profitable texts rich with dense intertextuality--a veritable goldmine for the comparativist. However flawed our method is, it is able to, at least, discern what ancient and modern scholarship has borne out; therefore, our method is on relatively sure footing. We have also demonstrated our method is able to detect similarities between fragments. Thus, we can use our method to classify unassigned fragments with some degree of confidence.

While our method has yielded some fruit there are many problems, questions or gaps that could be raised. These relate to deficiencies in my own knowledge as well as the method presented. Some of the problems in question could be corrected by either advances in the mathematical methods used, or in

building upon my research.

Most important among these gaps is that I am not a mathematician. While I am a Linux systems programmer by trade, I am completely self-taught in this area. The gaps in my knowledge of computer science and math are profound; therefore, I could have made simple blunders in the course of this dissertation that will be pointed out, no doubt, in the ensuing years.

Second, the list of algorithms that I presented and those that I eventually decided to use was not exhaustive. There exist many other data correlation algorithms that could have been used. Perhaps an algorithm was neglected that would have yielded better results or would have exposed other flaws in this dissertation. It is also possible that the perfect algorithm for document correlation could still yet be undiscovered.

In the use of our method above, SVD was used to process our document vectors without performing any normalization of the data beforehand (see page 45). While a normalization routine was written and tested before SVD was applied, it produced no appreciable difference in the results; therefore, it was excluded (see Appendix A). It is possible that a better normalization routine could have been used, but such a routine was unknown to me.

It is to be noted in our method that stop words, or common inconsequential words, were excluded from our documents (see page 44). These stop words are listed in Appendix C. Perhaps our results were skewed based upon words that should additionally have been excluded; or conversely, perhaps we excluded certain words that should have been included. I have listed these words in the Appendices for this reason.

Finally, another significant issue exists because of the way SVD works. You will remember after factoring our original matrix we are given three new matrices--Σ, V and U. The matrix that is refered to as Σ would give us a list of singular values (eigenvalues) from which the algorithm is named. These three matrices are multiplied together in order to give us a new matrix. This new matrix is then the matrix that we used to measure similarity. In some of the examples above, we had up to 36 singular values in Σ that we could either use or discard. In other words, we could create 35 different matrices from our factored matrix in this specific example. These new matrices sometimes differ greatly from the original matrix. Care was taken not to select a new matrix that diverged too greatly from the original matrix. Unfortunately, there is currently no automated way to know how many values should be retained or exluded from Σ (see page 40; Berry 54).

Further research could be done in the area of document correlation within the field of classics and comparative literature. Much could be done to build upon what has been done in this dissertation. Moreover, much could be done with document correlation within classics and comparative literature by going in different directions.

All our data has been listed in our Appendices. In addition, because the tools that I created and used are listed below in chapter 7, any data referred to in this dissertation can be duplicated. This data, at times, resulted in false-positives or false-negatives. That is, some of our data that has high coefficients may not be similar in content at all. Do the false-positives mean our entire method is invalid? Karl Pearson has taught us that invariably these anomalies in data will

occur. Care must be taken in order to verify our results. These coefficients are never gospel-truth similarities in documents, but serve as hints for us. In order to further validate or hone our method, some of these corrupt data could be examined more closely. Work could be done from a low-correlated poem to demonstrate that the poem in question is quite dense with intertextuality. This could show how our method is currently flawed and perhaps point out how this poem yielded such a skewed result in our method.

To further validate our classification of fragments an empirical test could be performed upon Horace, Juvenal and Persius. We could choose random lines with varying lengths from each author in an attempt to situate them into their respective poems. We could do this with these pseudo-fragments of specific lengths to determine if our method works well to classify fragments of a given length.

Further research could be done with more complex correlations in addition to our subject correlations, the proper name correlations and the lemma correlations. We could create a list of two or more words that have to appear in *n* number of lines. For example, we could use the words vir, homo, mulier and femina appearing in 2 or less lines. We would build our document vectors from any number of these rules in order to find areas that match our criteria. It is not hard to see how useful this would be for a classicist or a comparativist simply searching for a similar passage. Also, correlations could be performed using the scansion of specific lines. While metrical feet are standard, there is some variability within individual feet. Our document vector would then be built upon a certain metrical foot. For example, our correlation could be based upon finding

the following metrical paradigms (- stands for a long syllable where v stands for a short).

Table 6.1 Proposed metrical document correlation.

Dactyls	Document 1	Document 2
-vv -vv -- -- -- --	4	5
-- -- -- -- -vv --	8	7
-vv -vv -vv -vv -vv --	12	11
-- -vv -vv -- -vv --	3	0
-vv -- -vv -vv -vv --	10	15

As you can see in Table 6.1 we are looking for patterns within dactylic hexameter. We would build our document matrix from our poems and perform our similarity tests to see if they are closely correlated by meter. A last correlation that may not yield much fruit, but perhaps may be interesting nonetheless, would be a correlation based upon phonemic data. I have already built a correlation filter in the tool described below in chapter 7. This filter can be selected from the pull-down menu in the correlation tool in order to build a document vector that reduces words down to its phonemic values. For example, words in Latin-based alphabets can be reduced into the forms found below.

Table 6.2 Proposed phonemic document correlation.

Word Before Filter	Phonemically reduced
quis	PAS
hoc	QAP
potest	PAPASP
videre	VABALA
quis	PAS
potest	PAPASP
pati	PAPA

In Figure 6.2 quis is reduced to PAS where P stands for a voiceless stop (qu=k), A stands for a vowel (i) and P stands for a sibilant (s). We could get more detailed by indicating frontal vowels or back vowels or any number of linguistic attributes. Even though I created the linguistic filter during this dissertation I thought it subsequently tangential to my purpose. It may, however, prove useful to someone more interested in doing research in meter or prosody.

A final area of further research would be in the clustering of all classical works using k-means clustering. K-means clustering would allow us to visually represent the set of data points of our document vectors. Latin works that have some Greek words could cluster farther away from works that are purely Latin. For instance, Juvenal is primarily in Latin, but would quote a Greek hexameter, thus he could cluster farther away from Horace since Horace uses no Greek words, but since both wrote satire they would remain relatively close. Thus, we would expect the satirists to cluster together since they are in the same genre and at times write about the same subjects. We could even run the clustering upon individual poems or separate chapters of books. It would be interesting to see how all the classical authors cluster based upon lemma, especially if we did not separate authors by poetry or prose. Thus it would represent how authors cluster strictly by lemma words. This clustering might prove useful to group authors previously thought unrelated to one another.

Chapter 7 - Research Tools

Correlation Tool

I have developed a correlation tool during my research that could be useful for further research in Figure 7.1. It includes a variety of classical texts. You can use it to correlate any text against one or more other texts. Since neither SVD nor my method is language-specific you can correlate Greek texts as well. I have left this tool at the following url: http://beta.septuagint.org/correlate.

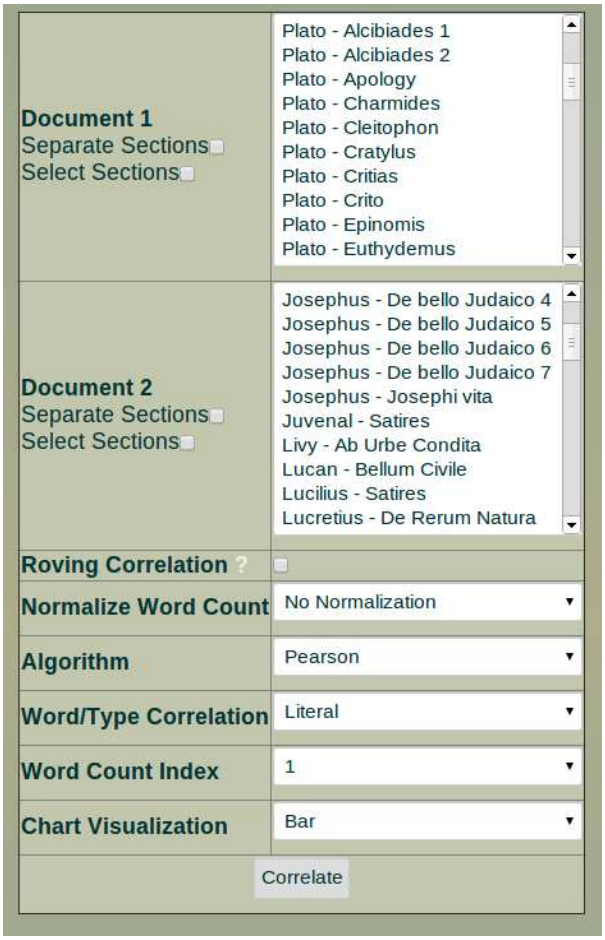


Figure 7.1 Document correlation tool.

Super Concordance

In order to augment my intertextual study I developed a concordance tool that can be used to find all occurrences of a particular word across all classical authors. You can search by lemma, a literal word that is morphologically marked, a group of words, the meanings of words to find significant semantically relevant passages, by tense or by case. This tool is incredibly powerful when trying to verify false-positives. It can also be used as a starting point to find texts that may yield interesting coefficients in the Correlation Tool. Or it can be used as simply a concordance for searching particular classical texts. You can find this tool at the following url: http://beta.septuagint.org/concordance.

Reading Tool

Last, it seemed natural, since I had to import these texts for the Super Concordance and the Correlation Tool, to create an online reader of texts. Unlike Perseus it is not sluggish and it is optimized for reading on handheld devices. Along with my dissertation I wanted to deliver tools that were useful for classicists to further research. The online reader can be found at this url: http://beta.septuagint.org.

REFERENCES

Adams, J. N. *The Latin Sexual Vocabulary.* Baltimore: The Johns Hopkins University Press, 1982. Print.

Aldrich, John. "Genuine and Spurious in Pearson and Yule." *Statistical Science,* Vol. 10, No. 4 (1995): 364-376. Print.

Alvarez-Lacalle, E, B. Dorow, J. P. Eckmann, E. Moses. "Hierarchical Structures Induce Long-Range Dynamical Correlations in Written Texts." *Proceedings of the National Academy of Sciences of the United States of America*, 103 (2006): 7956-7961. Print.

Alvo, Mayer, Kadir Ertas. "Graphical Methods for Ranking Data." *The Canadian Journal of Statistics*, 20 (1992): 469-482. Print.

Anderson, E., Z. Bai, C. Bischof, J. Demmel, J. Dongarra, J. D. Cruz, A. Greenbaum, S. Hammarling, A. McKenney, S. Ostrouchov, D. Sorensen. *LAPACK Users' Guide*. Philadelphia: Society for Industrial and Applied Mathematics, 1999. Print.

Anderson, William S. *Essays on Roman Satire.* Princeton: Princeton University Press, 1982. Print.

---. "The Programs of Juvenal's Later Books." *Classical Philology*, 57.3 (1962): 145-160. Print.

The Perseus Project. Ed. Gregory R. Crane. Mar. 1997. Dept. of Classics, Tufts U. 3 April 2011. <http://www.perseus.tufts.edu/>.

The Latin Library. Ed. William Carey. 28 April 2011. <http://www.thelatinlibrary.com/>.

Bailey, D. R. Shackleton. "Lights on Lucilius." *The Classical Journal,* 76.2 (1980): 117-118. Print.

Berry, M. W., Murray Brown. *Understanding Search Engines: Mathematical Modeling and Text Retrieval.* Philadelphia: Society for Industrial and Applied Mathematics, 1999. Print.

Berry, M. W., Z. Drmac, E. R. Jessup. "Matrices, Vector Spaces and Information Retrieval." *SIAM Review,* 41 (1999): 335-362. Print.

Berry, M., S. Dumais, G. O'Brien. "Using Linear Algebra for Intelligent Information Retrieval." *SIAM Review,* 37 (1995): 573-595. Print.

Berry, M., B. Hendrickson, P. Raghavan. "Sparse Matrix Reordering Schemes for Browsing Hypertext." Ed. J. Renegar, M. Shub, S. Smale. *Lectures in Applied Mathematics Vol. 32: The Mathematics of Numerical Analysis.* Providence: American Mathematical Society, 1996. Print.

Berry, M. W., R. D. Fierro. "Low-rank Orthogonal Decompositions for Information Retrieval Applications." *Numerical Linear Algebra with Applications, 3 (1996): 301-328. Print.*

Bodard, Gabriel, Daniel Paul O'Donnell. *Digital Medievalist.* 14 March 2008 GMT. University of Lethbridge. 17 March 2013. <http://www.digitalmedievalist.org/journal/4/DCeditorial/>.

Bond. Johannis. *Horatius Flaccus Cum Commentariis Selectissimis Variorum: & Scholiis Integris Johannis Bond.* Lugd(uni) Batavorum: Apud Franciscum Hackium., 1653. Print.

Boswell, John. *Christianity, Social Tolerance, and Homosexuality*. Chicago: The University of Chicago Press, 1981. Print.

Bramble, J. C. *Persius and the Programmatic Satire: A Study in Form and Image.* Cambridge: Cambridge University Press, 1974. Print.

Braund, S. H. *Greece and Rome: Roman Verse Satire.* Australia: Oxford University Press, 1992. Print.

Braund, Susanna Morton. *Juvenal Satires Book 1.* Cambridge: Cambridge University Press, 1996. Print.

Braund, Susanna Morton and Wendy Raschke. "Satiric Grotesques in Public and Private: Juvenal, Dr. Frankenstein, Raymond Chandler and 'Absolutely Fabulous." *Greece & Rome, Second Series*, 49.1 (2002): 62-84. Print.

Broder, Michael. *Mensura Incognita: Queer Kinship, Camp Aesthetics, and Juvenal's Ninth Satire.* Dissertation, The City University of New York. Ann Arbor: ProQuest/UMI, 2010. (Publication No. AAT 3426633.)

Chrol, E. Del. *Countercultural Responses to the Crisis of Masculinity in Late Republican Rome.* Dissertation, University of Southern California. Ann Arbor: ProQuest/UMI, 2006. (Publication No. AAT 3237762.)

Cicero. *On the Orator: Books I-2*. Trans. E. W. Sutton. Ed. H. Rackham. Cambridge: Harvard University Press, 1948. Print.

Cichorius, Conrad. *Untersuchungen zu Lucilius*. Berlin: Weidmannsche Buchhandlung, 1908. Print.

Colton, Robert E. "Ausonius and Juvenal." *The Classical Journal*, 69.1 (1973): 41-51. Print.

"Computer." *Wikipedia, The Free Encyclopedia.* 22 July 2004, 10:55 UTC. Wikimedia Foundation, Inc. 4 Dec. 2009. <http://en.wikipedia.org/wiki/Computer>.

Courtney, E. *A Commentary on the Satires of Juvenal.* London: Athlone Press, 1981. Print.

Crane, Greg. *Classics and the Computer: An End of the History.* Perseus Digital Library. Dept. of the Classics, Tufts U. Web. 4 Dec. 2009. <http://www.perseus.tufts.edu/~gcrane/blackwells.final.Crane.pdf>.

David, H. A. "Occurrence of Common Terms in Mathematical Statistics." *The American Statistician,* 49, No. 2 (1995): 121-133. Print.

Deerwester, S. Dumais, G. Furnas, T. Landauer, R. Harshman. "Indexing by Latent Semantic Analysis." *Journal of the American Society for Information Science, 41 (1990): 391-407. Print.*

Dominik, William J., William T. Wehrle. *Roman Verse Satire.* Wauconda: Bolchazy-Carducci Publishers, Inc., 1999. Print.

Duff, I., R. Grimes, J. Lewis. "Sparse Matrix Test Problems." *ACM Transactions on Mathematical Software,* 15 (1989): 1-14. Print.

Duff, J. Wight. *Roman Satire.* Berkeley: University of California Press, 1936. Print.

---. *Roman Satire: Its Outlook on Social Life (The University Series).* Berkeley: Archon Books, 1964. Print.

Eckart, C., G. Young. "The Approximation of One Matrix by Another of Lower Rank." *Psychometrika,* 1 (1936): 211-218. Print.

Edwards, Catharine. *The Politics of Immorality in Ancient Rome.* Cambridge: Cambridge University Press, 2002. Print.

Eilenberger, Sonja Jean. *The Development of Dialogue in the Novel: Wieland and Diderot.* Dissertation, University of Illinois. Ann Arbor: ProQuest/UMI, 1974. (Publication No. AAT 48106.)

Ellison, John W. *Nelson's Complete Concordance of the Revised Standard Version of the Bible.* New York: Nelson, 1957. Print.

Ernesti, August Wilhelm, and Gottfried Heinrich Schäfer. *Glossarium Livianum sive Index Latinitatis Exquisitoris.* Hildesheim: Gg Olms, 1966. Print.

Ferguson, John. *Juvenal: The Satires.* New York: Duckworth Publishers, 2003. Print.

Fiske, George Converse. *Lucilius and Horace.* Madison: University of Wisconsin, 1920. Print.

Fligner, Michael A., Joseph S. Verducci, Paul E. Blower. "A Modification of the Jaccard-Tanimoto Similarity Index for Diverse Selection of Chemical Compounds Using Binary Strings." *Technometrics*, 44, No. 2 (2002): 110-119. Print.

Forrest, Derek William. *Francis Galton: The Life and Work of a Victorian Genius.* London: Elec Books Ltd, 1974. Print.

Frakes, W. B., R. Baeza-Yates. *Information Retrieval: Data Structures & Algorithms.* Englewood Cliffs: Prentice-Hall, 1992. Print.

Frank, Tenney. "Horace's Description of a Scene in Lucilius." *The American Journal of Philology,* 46.1 (1925): 72-74. Print.

Freudenburg, Kirk. *Satires of Rome.* Cambridge: Cambridge University Press, 2001. Print.

---. *The Cambridge Companion to Roman Satire.* Cambridge: Cambridge University Press, 2005. Print.

---. *The walking Muse: Horace on the Theory of Satire.* Princeton: Princeton University Press, 1992. Print.

Fung, R., B. D. Favero. "Applying Bayesian Networks to Information Retrieval." *Communications of the ACM*, 58 (1995): 27-30. Print.

Gaisser, Julia Haig, T. Davina McClain. *Livy Book 1: Text (Bryn Mawr Latin Commentaries).* Oxford: Thomas Library, 2001. Print.

Gellar-Goad, T. H. M. *Lucretius' De Rerum Natura and Satire*. Dissertation, Chapel Hill. Ann Arbor: ProQuest/UMI, 2012. (Publication No. AAT 3512770.)

Glasbey, C. A. "A Reduced Rank Regression Model for Local Variation in Solar Radiation." *Journal of the Royal Statistical Society. Series C (Applied Statistics),* 41 (1992): 381-387. Print.

Glickman, Robert Jay, Gerrit Joseph Staalman. *Manual for the Printing of Literary Texts and Concordances by Computer.* Canada: University of Toronto Press, 1966. Print.

Gold, Barbara. *Literary Patronage in Greece and Rome.* Chapel Hill: The University of North Carolina Press, 2011. Print.

Good, I. J. "Some Applications of the Singular Decomposition of a Matrix." *Technometrics,* 11 (1969): 823-831. Print.

Gould, H.E., J. L. Whiteley. *Livy: Book 1.* London: Gerald Duckworth & Co. Ltd., 2004. Print.

Gowers, Emily. "Fragments of Autobiography in Horace Satires I." *Classical Antiquity,* 22.1 (2003): 55-91. Print.

---. *The Loaded Table: The Loaded Table: Representations of Food in Roman Literature.* Oxford: Oxford University Press, 1997. Print.

Greenaugh, J. B. *Livy, Books I and II*. New York: Caratzas Brothers, 1976. Print.

Harrington, James Matthew. *Mens Sana: Authorized Emotions and the Construction of Identity and Deviance in the Saturae of Juvenal.* Dissertation, University of Michigan. Ann Arbor: ProQuest/UMI, 2009. (Publication No. AAT 3382207.)

Henderson, John. "Persius' Didactic Satire: The Pupil as Teacher." *Ramus*, 20.1 (1991): 123-148. Print.

Highet, Gilbert. *The Anatomy of Satire.* Princeton: Princeton University Press, 1972. Print.

Hobson, M. P., Andrew H. Jaffe, Addrew R. Liddle, Pia Mukherjee, David Parkinson. *Bayesian Methods in Cosmology*. Cambridge: Cambridge University Press, 2010. Print.

Hooley, Daniel M. *Roman Satire.* Oxford: Blackwell Publishing, 2007. Print.

Horace. *Satires, Epistles, Ars Poetica.* Trans. H. R. Fairclough. Cambridge: Harvard University Press, 2005. Print.

Howard-Hill, T. H. *Literary Concordances.* Oxford: Pergamon Press, 1979. Print.

Hubert, Lawrence, Jacqueline Meulman, Willem Heiser. "Two Purposes for Matrix Factorization: A Historical Appraisal." *SIAM Review*, 42 (2000): 68-82. Print.

Hutchinson, G. O. *Latin Literature from Seneca to Juvenal.* Oxford: Clarendon Press, 1993. Print.

"Jaccard Coefficient." *Wikipedia, The Free Encyclopedia.* 22 July 2004, 10:55 UTC. Wikimedia Foundation, Inc. 30 Mar. 2013. <http://en.wikipedia.org/wiki/Jaccard_coefficient>.

Jefferis, J. D. "Juvenal and Religion." *The Classical Journal,* 34.4 (1939): 229-233. Print.

Jenkyns, Richard. "Juvenal's Satiric Persona." Rev. of "The Persona in Three Satires of Juvenal," by Martin M. Winkler. *The Classical Review, New Series,* 35.1 (1985): 34-36. Print.

Jones, Frederick. *Juvenal and the Satiric Genre.* London: Duckworth, 2007. Print.

Jope, James. "Lucian's Triumphant Cinaedus and Rogue Lovers." *Helios,* 36.1 (2009): 55-65. Print.

Juvenal, Persius. *Juvenal and Persius.* Trans. Susanna Morton Braund. Cambridge: Harvard University Press, 2004. Print.

---. *Juvenal: The Sixteen Satires.* Trans. Peter Green. London: Penguin Classics, 2004. Print.

---. *Sixteen Satires Upon the Ancient Harlot*. Trans. Steven Robinson. Manchester: Carcanet New Press, 1983. Print.

---. *Juvenal The* Satires. Trans. Niall Rudd, Ed. William Barr. Oxford: Oxford University Press, 1999. Print.

---. *D. Junii Juvenalis Saturarum Libri V.* Ed. Ludwig Friedlaender. Amsterdam: Verlag Adolf M. Hakkert, 1962. Print.

Keane, Catherine. *Figuring Genre in Roman Satire.* Oxford: Oxford University Press, 2006. Print.

---. "Philosophy Into Satire: The Program of Juvenal's Fifth Book." *American Journal of Philology*, 128.1 (2007): 27-57. Print.

---. "Satiric Memories: Autobiography and the Construction of Genre." *The Classical Journal*, 97.3 (2002): 215-231. Print.

Kemp, Jerome. "A Moral Purpose, A Literary Game: Horace, Satires 1.4." *Classical World,* 104.1 (2010): 59-76. Print.

Kiernan, V. G. *Horace: Poetics and Politics*. New York: St. Martin's Press, 1998. Print.

Knoche, Ulriche. *Roman Satire.* Trans. Edwin S. Ramage. Bloomington: Indiana University Press, 1975. Print.

Levin-Richardson, Sarah. *Roman Provocations: Interactions with Decorated Spaces in Early Imperial Rome and Pompeii.* Dissertation, Stanford University. Ann Arbor: ProQuest/UMI, 2009. (Publication No. AAT 3364133.)

Livy. *Ab Urbe Condita.* Perseus.org, Perseus Digital Library. Dept. of the Classics, Tufts U. Web. 4 Dec. 2009.

Long, Cliff. "Visualization of Matrix Singular Value Decomposition." *Mathematics Magazine*, 56 (1983): 161-167. Print.

Lucilius. *Remains of Old Latin.* Trans. E. H. Warmington. Ed. Jeffrey Henderson. Cambridge: Harvard University Press, 2004. Print.

Martyn, John R. C. "Juvenal and ne quid nimis." *Hermes,* 102.2 (1974): 338-345. Print.

Marx, Fredericus. *C. Lucilii Carminum Reliquiae Vol 1.* Leipzig: 1904. Print.

Marx, Fredericus. *C. Lucilii Carminum Reliquiae Vol 2.* Leipzig: 1904. Print.

Mason, Wyatt. "My Satirical Self." New York Times 17 Sep. 2006, E72.

Mayor, John E. B., Juvenal. *Thirteen satires of Juvenal Vol 1.* London: MacMillan and Company, 1886. Print.

Melville, Herman. *Moby-Dick.* New York: Barnes & Noble Classics, 2003. Print.

Miller, Paul Allen. "The Bodily Grotesque in Roman Satire: Images of Sterility." Arethusa 31.3 (1998): 257-283. *Project MUSE*. Web. 10 Nov. 2012. <http://muse.jhu.edu/>.

---. Latin Erotic Elegy. New York: Routledge Taylor & Francis Group, 2002. Print.

---. *Latin Verse Satire.* London: Routledge Taylor & Francis Group, 2005. Print.

Nash, J. C., L. P. Lefkovitch. "Principal Components and Regression by Singular Value Decomposition on a Small Computer." *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 25 (1976): 210-216. Print.

Oakman, Robert L. *Concordances from Computers ; A Review Article.* Columbia, S.C.: University of South Carolina University Press, 1973. Print.

Ogilvie, Robertus Maxwell. *Ab Urbe Condita: Volume 1: Books I-V.* Oxford: Oxford University Press Inc., 1974. Print.

Omberg, Larsson, Gene H. Golub, Orly Alter. "A Tensor Higher-Order Singular Value Decomposition for Integrative Analysis of DNA Microarray Data from Different Studies." *Proceedings of the National Academy of Sciences of the United States of America,* 104 (2007): 18371-18376. Print.

Osborne, Jason W. *Best Practices in Quantitative Methods*. Los Angeles: Sage Publications, Inc., 2008. Print.

Owen, Sean, Robin Anil, Ted Dunning, Ellen Friedman. *Mahout in Action.* Greenwich: Manning Publications Co., 2011. Print.

Packard, David W. *A Concordance to Livy: Vol I-IV.* Massachusetts: Harvard University Press, 1968. Print.

---. "Publishing Scholarly Compilations by Computer." *Computers and the Humanities,* Vol 4, Num 1(1969): 75-80. Print.

Paige, C. C., M. A. Saunders. "Towards a Generalized Singular Value Decomposition." *SIAM Journal on Numerical Analysis,* 18 (1981): 398-405. Print.

"Pearson Coefficient." *Wikipedia, The Free Encyclopedia.* 22 July 2004, 10:55 UTC. Wikimedia Foundation, Inc. 2 Mar. 2013. <http://en.wikipedia.org/wiki/Pearson_correlation_coefficient>.

Pereira, Maria Helena da Rocha and José Ribeiro Ferreira, Francisco de Oliveira. "Horácio e a sua perenidade." Coimbra: Instituto de Estudos Clássicos, 2009. Print.

Persius. *Persius: The* Satires. Trans. J. R. Jenkinson. Warminster: Aris & Phillips, 1981. Print.

Piwonka, Mario Puelma. *Lucilius und Kallimachos.* New York: Garland Publishing, Inc.: 1978. Print.

Plaza, Maria. *The function of Humour in Roman Verse Satire.* Oxford: Oxford University Press, 2006. Print.

Porter, Theodore M. *Karl Pearson.* Princeton: Princeton University Press, 2004. Print.

"Punchcard." *Wikipedia, The Free Encyclopedia.* 22 July 2004, 10:55 UTC. Wikimedia Foundation, Inc. 4 Dec. 2009. <http://en.wikipedia.org/wiki/Punchcard>.

Quintilian. *Institutio Oratoria.* Ed. Donald A. Russell. Cambridge: Harvard University Press, 2001. Print.

Richlin, Amy. *The Garden of Priapus: Sexuality and Aggression in Roman Humor.* New Haven: Oxford University Press, 1992. Print.

---. "Invective Against Women in Roman Satire." *Arethusa*, 17 (1984): 67-80. Print.

---. "Making Up a Woman: The Face of Roman Gender." Ed. Eilberg-Schwartz, Doniger, Wendy. *Off With Her Head*! *The Denial of Women's Identity in Myth*, *Religion*, *and Culture,* (1995): 185-213. Print.

Robinson, Laura. "The Personal Abuse in Lucilius' Satires." *The Classical Journal,* 49.1 (1953): 31-35. Print.

Sandberg, Rickard. "Assessment of Tumor Characteristic Gene Expression in Cell Lines Using a Tissue Similarity Index (TSI)." *Proceedings of the National Academy of Sciences of the United States of America,* 102 (2005): 2052-2057. Print.

Schlegel, Catherine. "Horace and His Fathers: Satires 1.4 and 1.6." *American Journal of Philology*, 121.1 (2000): 93-119. Print.

Scodel, Ruth. "Horace, Lucilius, and Callimachean Polemic." *Harvard Studies in Classical Philology,* 91 (1987): 199-215. Print.

Seeley, J. R. *Livy: Books I-X.* 2nd ed. Oxford: Clarendon Press. 1874. Print.

Taylor, Rabun. "Two Pathic Subcultures in Ancient Rome." *Journal of the History of Sexuality,* 7, 3 (1997): 319-371. Print.

Todeschini, Roberto, Viviana Consonni. *Molecular Descriptors for Chemoinformatics: Alphabetical Listing: Volume 1.* Weinheim: Wiley-VCH, 2009. Print.

Thelen, Ed. *BRL Report 1961. W*eb. 4 Dec. 2009. <http://www.ed-thelen.org/comp-hist/BRL61-ibm07.html>.

Ulden, James. *The Invisibility of Juvenal.* Dissertation, Columbia University. Ann Arbor: ProQuest/UMI, 2011. (Publication No. AAT 3453193.)

Umurhan, Osman Sami. *Spatial Representation in Juvenal's Satires: Rome and the Satirist.* Dissertation, New York University. Ann Arbor: ProQuest/UMI, 2008. (Publication No. AAT 3320849.)

---. "Poetic Projection in Juvenal's Satires." *Arethusa,* 44.2 (2011): 221-243. Print.

Van Loan, Charles F. "Generalizing the Singular Value Decomposition." *SIAM Journal on Numerical Anaysis,* 13 (1976): 76-83. Print.

Watt, W. S. "Notes on Juvenal." *Hermes,* 130, 3 (2002): 299-305. Print.

Weiss, Herold. "The Pagani among the Contemporaries of the First Christians." *Journal of Biblical Literature,* 86.1 (1967): 42-52. Print.

Williams, Craig A. *Roman Homosexuality*. Oxford: Oxford University Press, 1999. Print.

Winter, Thomas Nelson. "Roberto Busa, S.J., and The Invention of the Machine-Generated Concordance." *Classical Bulletin* 75 (1999): 3-21. Print.

Woods, Heather A. *Hunting Literary Legacies: Captatio in Roman Satire.* Dissertation, University of Minnesota. Ann Arbor: ProQuest/UMI, 2012. (Publication No. AAT 3549338.)

Worsley, Keith J. "Comparing Functional Connectivity via Thresholding Correlations and Singular Value Decomposition." *Philosophical Transactions: Biological Sciences,* 360 (2005): 913-920. Print.

Wyke, Maria. "Taking the Woman's Part: Engendering Roman Love Elegy." Roman Literature and Ideology: Ramus Essays for J. P. Sullivan. Ed. A. J. Boyle. Bendigo: Aureal Publications, 1995. Print.

Yeung, M. K. Stephen, Jesper Tegner, James J. Collins. "Reverse Engineering Gene Networks Using Singular Value Decomposition and Robust Regression." *Proceedings of the National Academy of Sciences of the United States of America,* 99 (2002): 6163-6168. Print.

Appendix A - Formulae

Euclidean Norm

$$\sqrt{x1^2 + x2^2 + x3^2 \ldots}$$

Euclidean Dot Product

$$(x1 \bullet y1) + (x2 \bullet y2) + (x3 \bullet y3) \ldots$$

Pearson

$$\frac{\Sigma x^n \bullet y^n - (\Sigma x^n \bullet \Sigma y^n / n)}{\sqrt{\Sigma x^{n2} - (\Sigma x^{n2} / n) \bullet \Sigma y^{n2} - (\Sigma y^{n2} / n)}}$$

Jaccard Similarity Coefficient

a = Total number where a particular word appears in both document 1 and 2
b = Total number where a particular word appears in document 2, but not 1
c = Total number where a particular word appears in document 1, but not 2
d = Total number where a particular word appears in neither document 1 nor 2

$$\frac{a}{b + c + a}$$

Jaccard Distance

a = Total number where a particular word appears in both document 1 and 2
b = Total number where a particular word appears in document 2, but not 1
c = Total number where a particular word appears in document 1, but not 2
d = Total number where a particular word appears in neither document 1 nor 2

$$\frac{b + c}{b + c + a}$$

Cosine Similarity

$$\frac{\Sigma x^n \bullet y^n}{\sqrt{\Sigma x^{n2}} \bullet \sqrt{\Sigma y^{n2}}}$$

Tanimoto Coefficient

$$\frac{\Sigma (x^n \wedge y^n)}{\Sigma (x^n \vee y^n)}$$

Spearman Coefficient or Spearman's ρ (rho)

$$\frac{\Sigma(x^n - \bar{x^n})(y^n - \bar{y^n})}{\sqrt{\Sigma(x^n - \bar{x^n})^2 \bullet (y^n - \bar{y^n})^2}}$$

How to Transpose a Matrix

A matrix is easily transposed by turning all rows into columns.

Original Matrix

1 1 1 1 1 1 1
2 2 2 2 2 2 2
3 3 3 3 3 3 3
4 4 4 4 4 4 4

Transposed Matrix
1 2 3 4
1 2 3 4
1 2 3 4
1 2 3 4
1 2 3 4
1 2 3 4
1 2 3 4

Original Matrix

```
0 0 0 0
0 0 0 0
0 M 0 0
T a h 0
h t a y
e r s o
0 i 0 u
0 x 0 0
0 0 0 0
0 0 0 0
```

Transposed Matrix

```
0 0 0 T h e 0 0 0 0
0 0 M a t r i x 0 0
0 0 0 h a s 0 0 0 0
0 0 0 0 y o u 0 0 0
```

PHP Normalization Algorithm

```
/* Given a variable $matrix that is a multi-dimensional array */
###################################
function normalize_matrix($matrix)
###################################
        {
        $new_matrix=array();
        $matrix=transpose_matrix($matrix);
        $cnt=0;
        foreach ($matrix as $vector)
                {
                $pnts=0;
                foreach ($vector as $pnt)
                        {
                        $pnts+=pow($pnt,2);
                        }
                $vl[$cnt]=sqrt($pnts);
                $cnt++;
                }
        $cnt=0;
        foreach ($matrix as $vector)
                {
                $new_vector=array();
                foreach ($vector as $pnt)
                        {
                        $new_pnt=sprintf("%.5f",($pnt / $vl[$cnt]));
                        array_push($new_vector,$new_pnt);
                        }
                $new_matrix[$cnt]=$new_vector;
                $cnt++;
                }
        return transpose_matrix($new_matrix);
        }
```

Appendix B - Word Lists for Subject Correlations

Animals

altilis, anguis, aratrum, aries, asinus, bos, bubulcus, caballus, canis, cantherius, catulus, cauda, cercurus, colubra, delphinus, echinus, elephantus, fera, fibra, ficedula, gallina, grus, helops, iugum, iumentum, leo, lustrum, mergus, mulus, murena, muscipulum, palumbes, pecus, pecus, peloris, pinna, pinnatus, pluma, polypus, porcus, purpura, rostrum, sargus, scorpius, sonipes, stabulum, sumen

The Body

sto/ma, anima, articulus, auricula, auris, barba, capillus, caput, caulis, cervix, cinerarius, clunis, collum, cor, corium, corpus, costa, coxa, crus, culus, dens, dextra, digitus, facies, fauces, folliculus, iecusculum, inguen, intercus, labrum, lacertus, lumbus, mamma, naevus, naris, nasus, nasutus, natis, nervus, oculus, os, os, palma, papilla, pectus, pedes, pellicula, penis, pes, planta, podex, posticus, praecordia, pulmo, rictus, sanguis, stomachus, sumen, sura, talus, tergus, testis, tonsillae, truncus, ulcus, unguis

Disease

aeger, aegritudo, aegrotus, amens, cicatrix, cludo, distentus, dolor, fames, fastidiosus, febris, frigus, gibbus, gravedo, horror, ictericus, incuria, insanus, lassus, lippus, macula, mancus, menda, morbus, mors, naevus, pallor, papula, porrigo, ruga, scabies, senex, senium, strabo, surdus, torpor, turdus, tussis, varicosus, varus, venenum, verruca, vescus, vetus, virus, vomica, vomitus, gangraena

The Dishonorable

caenum, calvus, carcer, caries, castro, cerebrosus, cinaedus, clepo, damnum, dolosus, elevo, exlex, exul, fama, famulus, fur, horridulus, humilis, idiota, ignavus, ignobilis, improbus, imprudens, impuratus, impurus, indignus, iners, infamis, infelix, infitiae, inhonestus, inimicus, iniuratus, inlitteratus, insanus, insidiae, insulsus, inutilis, ira, iratus, leno, limus, lucifugus, lustrum, lutum, macula, maculosus, maeror, malus, mastigia, mendicus, mendum, mentior, misellus, miser, moechus, molestus, nebulo, nefandus, nequam, nequitia, nugator, odiosus, odorus, pecco, periurus, petulantia, pigror, pinguesco, poena, propola, pudor, puer, quartarius, scelerosus, scurra, servitus, servus, sordidulus, spurcus, stercus, stulte, stultitia, stultus, superbus, surdus, taeter, taetre, tagax, tardus, torpor, tristis, tristitia, turpis, usura, verna, virus

Excess
amens, aurum, cachinnus, calix, centum, comedo, conficio, copia, daps, deliciae, devoro, distentus, divitiae, ebrius, elevo, gumia, gurges, iacio, indulgeo, irascor, lacrimosus, largus, lustrum, magnus, milia, mille, nummarius, nummus, omnis, pinguesco, pinguis, plenus, potus, sestertius, tantus, usura, ventriculus, vorax

Food List
a)ru/taina, allium, alo, altilis, alveolus, anser, asparagus, bibo, cadus, caleo, calix, carpo, caseus, catillus, catinus, cauda, cena, cenaculum, ceno, cepa, cibus, cocus, comedo, coquo, crustulum, culina, dominium, echinus, epulum, far, fervo, fibra, ficus, fructus, frumentarius, frumentum, gallinaceus, gallus, gusto, guttur, helops, herba, holus, hordeum, lanx, lardum, maena, mando, mappa, mensa, merum, molitus, molo, mordeo, murena, obsonium, oenophorum, oleum, omentum, ostrea, ovum, palumbes, panis, penus, piscis, pistrinum, placenta, popina, potus, potus, pulmentarium, puls, ructus, sal, sargus, seges, silurus, sodalicius, squilla, sumen, sumptus, urceolus, uva, ventriculus, vinum, viscus, viscera

The gods
Apollo, Camena, Ceres, deus, divinus, divus, dominus, fatum, fors, fortuna, Ianus, Iuppiter, mactus, Mars, Minerva, Musa, Neptunus, numen, omen, omnipotens, Orcus, sacrum, Saturnus, tus

Man & Virtue
a(mo/s, patria/, amator, amicus, argutus, bonus, caveo, consilium, cupide, dignus, doctus, dominus, facetus, fama, fautor, fides, formosus, fortiter, forum, gladiator, gymnasium, homo, honestus, honor, ingenium, iuventus, laus, legio, lex, libertas, lustratio, mortalis, munificus, munus, murus, nobilis, officium, parens, pater, pax, pietas, populus, praeclarus, praetor, primus, probatus, Quirinus, rectus, salus, salveo, sanus, sapiens, sapientia, sedulo, sedulus, servo, sophus, studiose, studiosus, urbs, utilis, verus, vir, virtus, vis, vita, vito

Proper Names
Acci, Accius, Acestes, Achaei, Achille, Achillem, Achilles, Achillis, Achivis, Acilius, Actiaca, Actoris, Aeacidae, Aeacus, Aegaei, Aegaeum, Aegypti, Aegyptius, Aegypto, Aegyptos, Aelia, Aemilianos, Aemilio, Aemilius, Aenea, Aeneae, Aenean, Aeneas, Aeoliis, Aeolio, Aeserninus, Aesopi, Aethiopem, Aethiopis, Aethiopum, Aetnae, Afra, Afrae, Africa, Afris, Afros, Agamemnona, Agamemnonidae, Agamemnonis, Aganippes, Agathyrsi, Agaue, Agauen, Agrion, Agrippa, Agrippinae, Aiacem, Aiax, Alabandis, Alba, Albana, Albanam, Albani, Albanis, Albanum, Albesia, Albinam, Albine, Albius, Albuci, Albucius, Alburnum, Alcestim, Alcinoo, Alcithoen, Alcmenam, Alcon, Alexander, Alfenus, Alledius, Allifanis, Allobroga, Allobrogicis, Alpem, Alpes, Alpibus, Alpinus, Alpis, Ambitio, Ambrosius, Amphion, Amphitryonis, Amyclas, Amydone, Ancarius, Anchemoli, Anchisae, Ancon, Andro, Andromachen, Andronis, Annales, Annibale, Annibalem, Antaeum, Anticatones, Anticyra, Anticyram, Anticyras, Antigones, Antilochi, Antiochus, Antiopa, Antiphates, Antoni, Antonius, Anubis, Anxur, Anyti, Aonidum,

Apella, Apelli, Apicius, Apollinis, Apollo, Apollost, Appennino, Appi, Appia, Appius, Apula, Apulia, Apuliam, Apulidae, Apulus, Aquarius, Aquilo, Aquino, Arabarches, Arabus, Arachne, Arbuscula, Arcadiae, Arcadico, Arcesilas, Archigene, Archigenen, Archilochum, Arciloco, Arelli, Argis, Aricia, Aricinos, Aristippum, Aristippus, Aristius, Aristocratem, Aristophanes, Aristotelen, Armeniae, Armenio, Armenius, Armillato, Arpinas, Arreti, Arri, Artaxata, Artemo, Artorius, Aruiragus, Asellus, Asiae, Asiam, Asiani, Asianorum, Assaraci, Assyrio, Astraea, Asyli, Atabulus, Atacino, Atellanae, Athenae, Athenas, Athenis, Athones, Athos, Atlanta, Atlas, Atreus, Atrida, Atriden, Atrides, Atridis, Attica, Atticon, Attis, Auaritia, Aufidio, Aufidius, Aufidus, Augusta, Augusto, Augustum, Aule, Aulide, ÄIulius, Aurelia, Auroram, Auruncae, Aurunci, Auster, Austri, Austris, Automedon, Autonoes, Aventini, Avidienus, Babylonem, Bacchae, Bacchanalia, Bacchius, Baeticus, Baianae, Baiano, Baiarum, Baias, Balatro, Balatrone, Balatroni, Balbinum, Baptae, Bardaicus, Baream, Bari, Barros, Barrus, Basilo, Basilum, Basilus, Bassaris, Basse, Bataui, Bathylli, Bathyllo, Baucis, Bebriaci, Belides, Bellerophonti, Bellona, Bellonae, Beneuentani, Beneventum, Berecyntius, Beronices, Bestius, Bibule, Birri, Bitho, Bithyni], Bithynice, Bithyno, Bitto, Blande, Boccare, Bolane, Bootae, Bouillas, Bovillanus, Brigantum, Brisaei, Britannica, Britannice, Britanno, Britannos, Brittones, Bromium, Brundisium, Brute, Bruti, Bruto, Brutorum, Bruttace, Bruttia, Bruttidius, Brutum, Brutus, Byzantia, Cacus, Cadmo, Caecuba, Caedicio, Caedicius, Caeli, Caesar, Caesare, Caesaris, Caesonia, Caetronius, Caietae, Calabrum, Calenum, Calliope, Callirhoen, Calpe, Calpurni, Caluine, Calvinae, Calvum, Camena, Camenae, Camenas, Camenis, Camerinos, Camerinus, Camilli, Campana, Campania, Campanis, Campano, Campanum, Campanus, Canem, Canicula, Canidia, Canidiae, Canidiam, Canis, Cannarum, Cannis, Canopi, Canopo, Cantaber, Canusi, Canusinam, Canusini, Capenam, Capito, Capitolia, Capitolinam, Capitolini, Capitolinis, Capitolinus, Capitone, Cappadocas, Cappadoces, Caprearum, Capreis, Capri, Caprius, Capua, Capuae, Carbo, Carfinia, Carneaden, Carpathium, Carpophoro, Carrinatis, Carthagine, Carus, Casinas, Cassandra, Cassandram, Cassi, Cassius, Castor, Castora, Castore, Cati, Catia, Catienae, Catienis, Catilina, Catilinam, Catinensi, Catius, Cato, Catone, Catonem, Catonis, Catuli, Catulla, Catullam, Catulli, Catullo, Catullum, Catullus, Catulus, Caudi, Cecilius, Cecropiam, Cecropides, Cecropis, Celaeno, Celso, Cephalonem, Cerco, Cererem, Cereris, Ceres, Cerinthe, Cervius, Cethegum, Cethegus, Chaerestratus, Chaerippe, Chaldaeis, Chaldaeo, Charybdi, Charybdim, Chattis, Chii, Chio, Chionen, Chiron, Chironeo, Chium, Chremes, Chremeta, Chrysidis, Chrysippe, Chrysippi, Chrysippus, Chrysogonum, Chrysogonus, Ciceronem, Ciceroni, Cicirri, Cicirrus, Cicutae, Cicutam, Cilicis, Cilicum, Cimbri, Cimbros, Circeis, Circes, Cirrhae, Cirrhaei, Claudius, Clazomenis, Cleanthas, Cleanthea, Cleopatra, Clio, Clitumni, Clodius, Cluuiam, Cluuienus, Clytemestram, Coa, Coccei, Cocceius, Coclite, Cocytum, Coelius, Cois, Colchide, Collina, Commagenus, Concordia, Congum, Coo, Copti, Corano, Coranum, Corbulo, Corcyraea, Cordi, Cordo, Cordus, Corinthi, Corinthon, Corneli, Cornelia, Cornelius, Cornifici, Cornute, Corsica, Coruine, Coruinum, Coruinus, Corum, Corvinus, Corybanta, Corycia, Corycio, Coryphaei, Cosmi, Cosso, Cossum, Cossus, Cotta, Cotus, Cotyton, Crassi, Crasso,

Crassos, Crassum, Cratero, Craterum, Cratino, Cratinus, Credo, Cremerae, Crepereius, Cressa, Cretae, Cretice, Creticus, Crispi, Crispine, Crispini, Crispinum, Crispinus, Croesi, Croesum, Crysi, Cumis, Cupiennius, Curibus, Curios, Curius, Curtillus, Curtius, Cyane, Cyaneis, Cybeles, Cyclada, Cyclopa, Cyclopas, Cyclops, Cycnum, Cynici, Cynicis, Cynicos, Cynthia, Dacicus, Dacis, Daedalus, Dama, Damae, Damasippe, Damasippi, Damasippus, Daue, Davo, Davum, Davus, Davusne, Decembri, Decii, Decinius, Decio, Deciorum, Decius, Decumum, Delphis, Delum, Demaenetus, Demetri, Demetrius, Democritus, Deucalion, Deucalione, Diana, Dianae, Dianam, Dicarchitum, Dinomaches, Diomede, Diomedeas, Diomedi, Dionysi, Diphilus, Discordia, ~Dolabella, Domiti, Domutionis, Dorica, Dorida, Druso, Drusorum, Drusus, Ecbatanam, Echion, Egeria, Egeriae, Eisocration, Electrae, Electran, Elissae, Elpenora, Emathii, Endymion, Enni, Ennosigaeum, Epicure, Epicuri, Epicurum, Epidaurius, Eponam, Eppia, Ergenna, Erinys, Eriphylae, Esquilias, Esquiliis, Etrusci, Etruscos, Etruscum, Euandri, Euandrum, Euganea, Eumenidum, Euphranoris, Euphraten, Eupolidem, Eupolin, Eupolis, Europen, Eurum, Euryalum, Fabii, Fabiis, Fabio, Fabios, Fabium, Fabius, Fabrateriae, Fabricio, Fabricium, Fabricius, Fabulla, Facelinae, Faesidium, Falerna, Falernas, Falerni, Falerno, Falernum, Fanni, Fannius, Fauni, Fausta, Fausti, Feronia, Fidenarum, Fidenis, Fides, Flacci, Flaccorum, Flaccus, Flaminia, Flaminiam, Flavi, Flora, Florae, Florali, Floralia, Fonteio, Fonteius, fora, foro, fortuna, Fortuna, Fortunae, Forum, Frontonis, Frusinone, Fufidius, Fufius, Fulvi, Fundani, Furiae, Furiam, Furiis, Furius, Furni, Fusci, Fuscine, Fusco, Fuscus, Gabba, Gabiis, Gabiorum, Gadibus, Gaditana, Gaetula, Gaetulice, Gaetulum, Gaetulus, Gai, Gaius, Galba, Galbam, Galla, Galli, Gallia, Gallicus, Gallina, Gallinaria, Gallis, Gallitta, Gallittae, Galloni, Gallos, Gallus, Gangen, Ganymedem, Gargonius, Gaurana, Gaurus, Gemino, Geminos, Gentius, Germanae, Germani, Germanicus, Geticis, Gillo, Glaphyrus, Glauco, Gloria, Glyconi, Gnatho, Gnatia, Gorgone, Gorgonei, Gracchi, Graccho, Gracchorum, Gracchos, Gracchum, Gracchus, Gradiue, Gradiuus, Graecam, Graece, Graeci, Graecia, Graecis, Graecorum, Graecos, Graecula, Graeculus, Graecum, Graecus, Graiae, Graias, Graiorum, Graios, Graius, Grani, Granius, Gurgitis, Gyarae, Gyaris, Hadriaci, Haemo, Hagnae, Hamillus, Hammonis, Hannibal, Hannibalem, Hannibali, Harpyiis, Hecaten, Hectora, Hectore, Hedymeles, Helenam, Heliadum, Helicone, Heliconidas, Heliodorus, Hellade, Heluidius, Helvinam, Heracleas, hercle, Hercule, Herculeo, Herculeos, Herculis, Hermae, Hermarchus, Hermogenes, Hermogenis, Hernicus, Herodis, Hibera, Hiberi, Hiberinae, Hippolyto, Hirrus, Hispania, Hispo, Hispulla, Hister, Histro, Histrum, Homeri, Homericus, Homero, Homerum, Horatius, Hortensi, Hortensius, Hostilius, Hyacintho, Hyacinthos, Hydaspes, Hylas, Hymettia, Hymetto, Hymnidis, Hymnis, Hyperboreum, Hypsaea, Hypsipylas, Ianum, Ianus, Icadion, Idymaeae, Ilias, Illyricum, Isiacae, Isidis, Italo, Ithacum, Ithacus, Iunonem, Iuppiter, Ixionies, Karthagine, Labeone, Labeonem, Laberi, Lacedaemonium, Lacertae, Lachesi, Lachesis, Ladas, Laeli, Laelium, Laelius, Laenas, Laertiade, Laestrygonas, Laevino, Laevinum, Laevius, Lagi, Lamia, Lamiarum, Lamias, Laomedontiades, Lappa, Lare, Larem, Lares, Largae, Laribus, Laronia, Lateranorum, Lateranus, Latiis, Latina, Latinae, Latine, Latine.], Latini, Latino, Latio, Latona, Latonae, Lauino, Laurens, Laurenti, Laureolum, Lavernae, Ledae,

Ledam, Lentule, Lentulus, leontado, Lepidi, Lepidis, Lepos, Leucade, Liber, Libitinae, Libitinam, Libra, Liburna, Liburno, Liburnus, Libya, Licini, Licinis, Licinius, Licinus, Lictores, Ligus, Ligustica, Liparaea, Liparas, Longarenus, Longinum, lovis, Lucana, Lucanis, Lucanos, Lucanus, Luci, Lucili, Lucilio, Lucilium, Lucilius, Lucius, Lucretia, Lucrina, Lucrinum, Lucusta, Lugudunensem, Luna, Lunai, lunium, Lupe, luperco, Lupo, luppiter, Lupus, Lusco, Luxuria, Lycio, Lyciscae, Lycius, Lydorum, Lymphis, Lyncei, Lysippi, Macedo, Machaerae, Macrine, Maecenas, Maecenate, Maecenatem, Maecenatibus, Maedos, Maenas, Maenius, Maeonides, Maeotica, Maeotide, Maia, Maltinus, Mamercorum, Mamurrarum, Manil>ium, Manilia, Manius, Manlius, Man<ium, Marce, Marcellis, Marco, Marcus, Marius, Maronem, Maroni, Maronis, Mars, Marsaeus, Marsi, Marsos, Marsus, Marsya, Marti, Martis, Massa, Massica, Masuri, Matho, Mathonis, Matutine, Maura, Maurae, Mauri, Mauro, Maurorum, Maurus, Maximus, Medis, Medo, Medullinae, Megalesia, Megalesiacae, Melanippes, Meleagri, Melicerta, Memnona, Memnone, Memphitide, Menandro, Menelaum, Meneni, Menoeceus, Mentore, Mercuriale, Mercurialem, Mercurium, Mercurius, Meroe, Messalae, Messalinae, Messalla, Messanam, Messi, Messius, Metellae, Metelli, Metello, Metellorum, Methymnaeam, Metrophanes, Meuia, Micipsarum, Miletos, Miloni, Milonius, Mimalloneis, Mineruae, Minerva, Minervae, Minervam, Minturnarum, Miseno, Mithridates, Modiam, Moesorum, Molossis, Molosso, Molossos, Montani, Montanus, Monychus, Moyses, Muci, Mucius, Murena, Musa, Musae, Musarum, Musas, Musconis, Mycale, Mycenis, Myconi, Myronis, Nabataeo, Naeuole, Naevius, Narcissi, Nasica, Nasicae, Nasidiene, Nasidieni, Nasidienus, Natta, Nattae, Neptune, Neptuni, Neptunus, Nerea, Nerei, Nerio, Nero, Nerone, Neronem, Neroni, Neronis, Nestora, Nestoris, Nili, Niliacae, Nilo, Nilum, Niobe, Niphaten, Nomentane, Nomentani, Nomentano, Nomentanum, Nomentanus, Nortia, Nostius, Nouium, Noviorum, Novium, Novius, Numa, Numae, Numantinos, Numeri, Numidarum, Numidas, Numitor, Nysae, Oceani, Oceanum, Octauius, Octavius, Ofelli, Ofello, Ofellum, Ofellus, Olynthi, Ombis, Ombos, Opimius, Oppia, Oppidius, Orbiliae, Orcadas, Orco, Orcus, Oreste, Orestes, Originis, Orontes, Osci, Osiri, Osiris, Ostia, Othoni, Othonis, Oufente, Oufentina, Pacci, Paceni, Pacideiani, Pacideiano, Pacideianum, Pacilius, Pacis, Pacius, Pactolus, Pacuuio, Pacuuium, Pacuuius, Pacuviano, Paean, Palaemon, Palaemonis, Palantino, Palati, Palatia, Palatino, Palfurio, Palilia, Palinurum, Pallante, Pamphilum, Pansa, Pantilius, Pantolabo, Pantolabum, Papiria, Parcae, Paridem, Paridi, Paris, Parnaso, Parrhasii, Parthenio, Parthi, Parthis, Partho, Pauli, Paulo, Paulus, Pausiaca, Pavus, Pax, Pecunia, Pediatia, Pedio, Pedius, Pedo, Pegaseium, Pegasus, Pelea, Peleus, Pellaeo, Pelopea, Penatis, Penelopam, Penelope, Perelli, Pergula, Peribomius, Pericli, Persi, Persica, Persice, Persicus, Persium, Persius, Petilli, Petosiris, Phaeaca, Phaeacum, Phalarim, Phalaris, Phario, Pharon, Phasma, Phialen, Phidiacum, Philippi, Philippica, Philodemus, Philomela, Phoebi, Pholo, Phrygia, Phrygibus, Phrygio, Phryne, Phryx, Phyllidas, Picenis, Picens, Pico, Pieria, Pierides, Pierio, Pirenen, Pisaeae, Piso, Pisonis, Pitholeonti, Pittacon, Platona, Plotius, Pluton, Poeno, Polemon, Pollio, Pollittas, Polycliti, Polydamas, Polyphemi, Polyphemus, Polyxena, Pompeio, Pompeios, Pompeius, Pompilii, Pomponius, Pomptina, Ponti, Pontia, Pontica, Pontice, Ponticus, Ponto, Popili, Poplicola, Poppaeana,

Porcius, Postume, Postumius, Praenestinis, Praenestinus, Priami, Priamiden, Priamus, Priapi, Priapum, Priscus, Priverno, Prochytam, Procne, Procnes, Procula, Proculas, Proculeius, Promethea, Prometheus, Proserpina, Protogenes, Psecas, Publi, Publius, Pudicitiae, Pudicitiam, Pulfenius, Punica, Pusillam, Puteal, Pygmaea, Pygmaeus, Pyladen, Pylades, Pylius, Pyrenaeum, Pyrgensia, Pyrrha, Pyrrhum, Pythagorae, Pythagoran, Pythagoras, Pythagoreis, Pythagoreo, Pythia, Quinte, Quinti, Quintiliane, Quintiliano, Quintilianus, Quintillae, Quintus, Quirine, Quirini, Quirinos, Quirinus, Quiritem, Quirites, Rauola, Regina, regis, Remi, Remus, Rex, Rhadamanthus, Rheni, Rheno, Rhenos, Rhodi, Rhodio, Rhodios, Rhodopes, Rhodum, Rhodus, Rhondes, Roma, Romae, Romam, Romana, Romanam, Romane, Romanis, Romano, Romanorum, Romanum, Romanus, Romule, Romuleae, Romulidae, Roscius, Rubos, Rubrenus, Rubrius, Rufam, Rufillus, Rufum, Rufus, Rupili, Rusonem, Rutilae, Rutilo, Rutilus, Rutubae, Rutulis, Rutulum, Rutupinoue, Sabella, Sabellam, Sabellis, Sabina, Sabino, Sabinos, Sabinum, Sagana, Saganae, Saguntina, Salamine, Saleiio, Salernam, Sallustius, Sameramis, Samia, Samiam, Samio, Samnis, Samo, Samothracum, Santonico, Sardanapalli, Sardiniensem, Sardus, Sarmata, Sarmenti, Sarmentus, Sarrana, Satureiano, Saturnalibus, Saturni, Saturnia, Saturno, Saturnum, Saturnus, Satyrum, Saufeia, Sauromatae, Sauromatas, Scaevae, Scantinia, Scaurorum, Scauros, Scipiadae, Scipiadam, Scipiadas, Scyllam, Scythicae, Secundi, Seiano, Seianum, Seianus, Seiio, Seleuco, Seneca, Senecae, Senecam, Senonum, Septembri, Septembris, Seres, Sergiolus, Sergius, Seripho, Serrano, Seruilia, Servi, Servilio, Servius, Setinum, Sexte, Sibyllae, Sicula, Siculi, Siculo, Siculos, Sicyone, Sicyonia, Signinum, Silanus, Silari, Siluano, Sinuessae, Siren, Sirena, Sisennas, Sisyphus, Socratici, Socratico, Socraticos, Socraticum, Solis, Solones, Solymarum, Sophocleo, Sostratus, Spartana, Spartani, Spartano, Staberi, Staio, Statius, Stentora, Stertinius, Stheneboea, Stoica, Stoice, Stoicidae, Stoicus, Stratocles, Stygio, Subura, Suburae, Sulgi, Sulgius, Sullae, Sulmonensi, superbos, Superbus, Sura, Surrentina, Surrentinum, Sybaris, Syenes, Sygambris, Symmacus, Syphacem, Syra, Syracusis, Syri, Syriae, Syrium, Syro, Syrophoenix, Syrorum, Syrus, Tadius, Tagi, Tagus, Tanain, Tanaquil, Tantalus, Tappulam, Tarento, Tarentum, Tarpa, Tarpeia, Tarpeio, Tarpeium, Tarquinius, Tatio, Taurica, Tauromenitanae, Tedia, Telamonem, Telephus, Telesine, Tentura, Terea, Terenti, Terentiae, Teresian, Terpsichoren, Terrae, Tessalam, Teucrorum, Teucrum, Teutonico, Thabraca, Thaida, Thais, Thaletis, Tharsimachi, Thebaidos, Thebarum, Thebas, Thebe, Thebis, Themison, Theodori, Thersitae, Thersites, Theseide, Thessaliae, Thestiados, Thraces, Thracum, Thraex, Thrasea, Thrasylli, Thrax, Thurinus, Thyestae, Thyle, Thymele, Thymeles, Tiberi, Tiberim, Tiberino, Tiberinum, Tiberinus, Tiburis, Tiburte, Tiburtia, Tiburtino, tierei, Tigelli, Tigellius, Tigillinum, Tilli, Tiresia, Tiresiai, Tiresias, Tirynthius, Tisiphone, Tisiphonen, Titan, Titanida, Tite, Titio, Titos, Tityi, Tonantem?], Tongilii, Trallibus, Trausius, Trebati, Trebellius, Trebio, Trebium, Trebius, Treboni, Trifolinus, Triphallo, Triquetra, Tritani, Trivici, Troginus, Troia, Troiades, Troiae, Troianum, Troica, Troiugenae, Troiugenas, Troiugenis, Trypheri, Tubulus, Tuccia, Tuditanus, Tulli, Tullia, Tullius, Turbonis, Turius, Turni, Turnus, Tusca, Tusci, Tuscis, Tusco, Tusculidarum, Tuscum, Tutor, Tydides, Tyndaridarum, Tyndaris, Tyrias, Tyrio, Tyrius, Tyrrhenam, Tyrrhenos,

Tyrrhenum, Ucalegon, Ulixen, Ulixes, Umbreni, Ummidius, Vagelli, Valeri, Valgius, Varillus, Varium, Varius, Varrone, Vascones, Vaticano, Veiientanum, Veiiento, Velina, Venafranae, Venafrano, Venafri, Veneri, Veneris, Veneto, Ventidio, Ventidius, Venus, Venusina, Venustinam, Vergilio, Vergilium, Vergilius, Verginia, Verginius, Verrem, Verres, Verri, Vestam, Vestinus, Vibidius, Victoria, Villius, Vindice, Virbi, Viriato, Virro, Virroni, Virronibus, Virronis, Virtus, Viscorum, Viscum, Viscus, Viselli, Vlixes, Vltor, Vlubris, Vmbris, Volanerius, Volcania, Volcano, Volcanus, Volesos, Volscorum, Volsiniis, Volturnus, Volusi, Voranus, Vortumnis, Vrsidio, Vulcani, Vulcaniam, Zacynthos, Zalaces, Zenonis, Zopyriatim, Zopyrion, ᾽Αρες, χῖός

Speech
ambages, aruspex, bilinguis, blanditia, carmen, clandestinus, communico, dico, doceo, doctrina, doctus, eloquium, epistula, inlitteratus, laudo, lego, littera, loquor, maledico, modus, monogrammos, muttio, nefandus, numerus, oratio, poema, rhetoricoteros, scribo, scriptor, sententia, sermo, sophistes, taceo, verbum, versus, vocabulum

War Language
accido, anceps, ancile, arma, armamenta, ballista, bellum, castra, catapulta, centurio, cingo, clamo, classis, depugno, dominium, exercitus, ferrum, ferveo, gladiator, gladius, hasta, hostis, incitus, insidiae, interficio, internecio, invado, iter, labor, mereo, miles, navis, palaestra, paludatus, pellis, pila, plaga, praesidium, proeliator, proelior, proelium, pugna, pugno, remus, rorarii, sarisa, scutum, signifer, socius, sparus, tela, tragula, velox, vinco

Women
a)ndro/gunos, amica, amo, ancilla, androgynus, anus, caupona, cognata, conciliatrix, domina, domus, femina, forma, gnata, honestas, illa-, impuratus, intus, lacto, lanificus, liber, lupa, mamma, mater, medica, mulier, nupta, papilla, pulcher, redimiculum, saga, scortator, soror, sumen, tela, textor, torus, uxor, verro, virgo

Appendix C - Stop words

Latin Stop Words (from Perseus.org)

ab, ac, ad, adhic, aliqui, aliquis, an, ante, apud, at, atque, aut, autem, cum, cur, de, deinde, dum, ego, enim, ergo, es, est, et, etiam, etsi, ex, fio, haud, hic, iam, idem, igitur, ille, in, infra, inter, interim, ipse, is, ita, magis, modo, mox, nam, ne, nec, necque, neque, nisi, non, nos, o, ob, per, possum, post, pro, quae, quam, quare, qui, quia, quicumque, quidem, quilibet, quis, quisnam, quisquam, quisque, quisquis, quo, quoniam, sed, si, sic, sive, sub, sui, sum, super, suus, tam, tamen, trans, tu, tum, ubi, uel, uero

Greek Stop Words (from Perseus.org)

μή, ἑαυτοῦ, ἄν, ἀλλ', ἀλλά, ἄλλοσ, ἀπό, ἄρα, αὐτόσ, δ', δέ, δή, διά, δαί, δαίσ, ἔτι, ἐγώ, ἐκ, ἐμόσ, ἐν, ἐπί, εἰ, εἰμί, εἶμι, εἰσ, γάρ, γε, γα^, ἡ, ἥ, καί, κατά, μέν, μετά, μή, ὁ, ὅδε, ὅσ, ὅστισ, ὅτι, οὕτωσ, οὗτοσ, οὔτε, οὖν, οὐδείσ, οἱ, οὐ, οὐδέ, οὐκ, περί, πρόσ, σύ, σύν, τά, τε, τήν, τῆσ, τῇ, τι, τί, τισ, τίσ, τό, τοί, τοιοῦτοσ, τόν, τούσ, τοῦ, τῶν, τῷ, ὑμόσ, ὑπέρ, ὑπό, ὡσ, ὦ, ὥστε, ἐάν, παρά, σόσ

Moby Dick Stop Words

a,able,about,across,after,all,almost,also,am,among,an,and,any,are,as,at,be, because,been,but,by,can,cannot,could,dear,did,do,does,either,else,ever,every, for,from,get,got,had,has,have,he,her,hers,him,his,how,however,i,if,in,into,is,it,its, just,least,let,like,likely,may,me,might,most,must,my,neither,no,nor,not,of,off,often, on,only,or,other,our,own,rather,really,said,say,says,she,should,since,so,some, than,that,the,their,them,then,there,these,they,this,tis,to,too,twas,us,very,wants, was,we,were,what,when,where,which,while,who,whom,why,will,with,would,yet, you,your

APPENDIX D - All coefficients of H, P and J against L

Book 1 1 - 52		
Poems	**Poem Length**	**Coefficient**
Lucilius - Satires	188	1.00000
Juvenal - Satires - 1	770	0.80812
Juvenal - Satires - 2	737	0.84899
Juvenal - Satires - 3	1419	0.83594
Juvenal - Satires - 4	671	0.85941
Juvenal - Satires - 5	752	0.93531
Juvenal - Satires - 6	3085	0.70438
Juvenal - Satires - 7	1096	0.89096
Juvenal - Satires - 8	1169	0.93067
Juvenal - Satires - 9	696	0.96081
Juvenal - Satires - 10	1689	0.70360
Juvenal - Satires - 11	946	0.80970
Juvenal - Satires - 12	571	0.77026
Juvenal - Satires - 13	1162	0.82845
Juvenal - Satires - 14	1524	0.85280
Juvenal - Satires - 15	811	0.72318
Juvenal - Satires - 16	267	0.92742
Persius - Satires - Prologus	46	0.87984
Persius - Satires - 1	619	0.97328
Persius - Satires - 2	369	0.92721
Persius - Satires - 3	573	0.97322
Persius - Satires - 4	234	0.96524
Persius - Satires - 5	923	0.98381
Persius - Satires - 6	381	0.97045
Horace - Satires - 1.1	631	0.99733

Horace - Satires - 1.2	694	0.98827
Horace - Satires - 1.3	706	0.99544
Horace - Satires - 1.4	735	0.99644
Horace - Satires - 1.5	499	0.93668
Horace - Satires - 1.6	679	0.93792
Horace - Satires - 1.7	164	0.96133
Horace - Satires - 1.8	239	0.93735
Horace - Satires - 1.9	414	0.97627
Horace - Satires - 1.10	477	0.99361
Horace - Satires - 2.1	421	0.99775
Horace - Satires - 2.2	694	0.99267
Horace - Satires - 2.3	1657	0.89424
Horace - Satires - 2.4	448	0.97424
Horace - Satires - 2.5	568	0.98916
Horace - Satires - 2.6	612	0.99749
Horace - Satires - 2.7	592	0.95465
Horace - Satires - 2.8	461	0.99642

Book 2 53 - 93		
Poems	**Poem Length**	**Coefficient**
Lucilius - Satires	129	1.00000
Juvenal - Satires - 1	770	0.87245
Juvenal - Satires - 2	737	0.89906
Juvenal - Satires - 3	1419	0.84249
Juvenal - Satires - 4	671	0.90119
Juvenal - Satires - 5	752	0.93594
Juvenal - Satires - 6	3085	0.78244
Juvenal - Satires - 7	1096	0.93414
Juvenal - Satires - 8	1169	0.94452
Juvenal - Satires - 9	696	0.96012
Juvenal - Satires - 10	1689	0.76594
Juvenal - Satires - 11	946	0.85300
Juvenal - Satires - 12	571	0.83089
Juvenal - Satires - 13	1162	0.87461

Juvenal - Satires - 14	1524	0.89181
Juvenal - Satires - 15	811	0.79054
Juvenal - Satires - 16	267	0.93451
Persius - Satires - Prologus	46	0.88344
Persius - Satires - 1	619	0.97603
Persius - Satires - 2	369	0.96784
Persius - Satires - 3	573	0.98290
Persius - Satires - 4	234	0.97740
Persius - Satires - 5	923	0.99224
Persius - Satires - 6	381	0.92737
Horace - Satires - 1.1	631	0.98908
Horace - Satires - 1.2	694	0.99660
Horace - Satires - 1.3	706	0.97592
Horace - Satires - 1.4	735	0.97801
Horace - Satires - 1.5	499	0.94055
Horace - Satires - 1.6	679	0.88231
Horace - Satires - 1.7	164	0.94961
Horace - Satires - 1.8	239	0.95696
Horace - Satires - 1.9	414	0.93946
Horace - Satires - 1.10	477	0.98275
Horace - Satires - 2.1	421	0.98205
Horace - Satires - 2.2	694	0.98143
Horace - Satires - 2.3	1657	0.91211
Horace - Satires - 2.4	448	0.95860
Horace - Satires - 2.5	568	0.96064
Horace - Satires - 2.6	612	0.98267
Horace - Satires - 2.7	592	0.92429
Horace - Satires - 2.8	461	0.98086

Book 3 94 - 148		
Poems	**Poem Length**	**Coefficient**
Lucilius - Satires	185	1.00000
Juvenal - Satires - 1	770	0.94284
Juvenal - Satires - 2	737	0.96304

Juvenal - Satires - 3	1419	0.91606
Juvenal - Satires - 4	671	0.96415
Juvenal - Satires - 5	752	0.95851
Juvenal - Satires - 6	3085	0.85212
Juvenal - Satires - 7	1096	0.98311
Juvenal - Satires - 8	1169	0.98271
Juvenal - Satires - 9	696	0.96388
Juvenal - Satires - 10	1689	0.83186
Juvenal - Satires - 11	946	0.93848
Juvenal - Satires - 12	571	0.92123
Juvenal - Satires - 13	1162	0.94608
Juvenal - Satires - 14	1524	0.96246
Juvenal - Satires - 15	811	0.88272
Juvenal - Satires - 16	267	0.96882
Persius - Satires - Prologus	46	0.93613
Persius - Satires - 1	619	0.95081
Persius - Satires - 2	369	0.96439
Persius - Satires - 3	573	0.96642
Persius - Satires - 4	234	0.95460
Persius - Satires - 5	923	0.95868
Persius - Satires - 6	381	0.87652
Horace - Satires - 1.1	631	0.96143
Horace - Satires - 1.2	694	0.98856
Horace - Satires - 1.3	706	0.94805
Horace - Satires - 1.4	735	0.94805
Horace - Satires - 1.5	499	0.97533
Horace - Satires - 1.6	679	0.83992
Horace - Satires - 1.7	164	0.95020
Horace - Satires - 1.8	239	0.98504
Horace - Satires - 1.9	414	0.88801
Horace - Satires - 1.10	477	0.96173
Horace - Satires - 2.1	421	0.94286
Horace - Satires - 2.2	694	0.96532
Horace - Satires - 2.3	1657	0.83094
Horace - Satires - 2.4	448	0.96070
Horace - Satires - 2.5	568	0.91434

Horace - Satires - 2.6	612	0.95217
Horace - Satires - 2.7	592	0.88012
Horace - Satires - 2.8	461	0.94467

Book 4 149 - 185		
Poems	**Poem Length**	**Coefficient**
Lucilius - Satires	155	1.00000
Juvenal - Satires - 1	770	0.96669
Juvenal - Satires - 2	737	0.96446
Juvenal - Satires - 3	1419	0.88484
Juvenal - Satires - 4	671	0.97559
Juvenal - Satires - 5	752	0.91827
Juvenal - Satires - 6	3085	0.86513
Juvenal - Satires - 7	1096	0.99189
Juvenal - Satires - 8	1169	0.95894
Juvenal - Satires - 9	696	0.92538
Juvenal - Satires - 10	1689	0.88268
Juvenal - Satires - 11	946	0.93938
Juvenal - Satires - 12	571	0.93961
Juvenal - Satires - 13	1162	0.96220
Juvenal - Satires - 14	1524	0.96475
Juvenal - Satires - 15	811	0.91749
Juvenal - Satires - 16	267	0.95700
Persius - Satires - Prologus	46	0.92016
Persius - Satires - 1	619	0.91914
Persius - Satires - 2	369	0.96548
Persius - Satires - 3	573	0.94052
Persius - Satires - 4	234	0.93073
Persius - Satires - 5	923	0.94013
Persius - Satires - 6	381	0.82202
Horace - Satires - 1.1	631	0.93916
Horace - Satires - 1.2	694	0.96988
Horace - Satires - 1.3	706	0.91351
Horace - Satires - 1.4	735	0.91158

Horace - Satires - 1.5	499	0.94736
Horace - Satires - 1.6	679	0.77547
Horace - Satires - 1.7	164	0.92984
Horace - Satires - 1.8	239	0.98341
Horace - Satires - 1.9	414	0.83594
Horace - Satires - 1.10	477	0.92640
Horace - Satires - 2.1	421	0.91548
Horace - Satires - 2.2	694	0.93031
Horace - Satires - 2.3	1657	0.83741
Horace - Satires - 2.4	448	0.92404
Horace - Satires - 2.5	568	0.87014
Horace - Satires - 2.6	612	0.91717
Horace - Satires - 2.7	592	0.82266
Horace - Satires - 2.8	461	0.91855

Book 5 186 - 251		
Poems	**Poem Length**	**Coefficient**
Lucilius - Satires	246	1.00000
Juvenal - Satires - 1	770	0.87223
Juvenal - Satires - 2	737	0.88929
Juvenal - Satires - 3	1419	0.92768
Juvenal - Satires - 4	671	0.93853
Juvenal - Satires - 5	752	0.93272
Juvenal - Satires - 6	3085	0.70058
Juvenal - Satires - 7	1096	0.93750
Juvenal - Satires - 8	1169	0.96161
Juvenal - Satires - 9	696	0.93201
Juvenal - Satires - 10	1689	0.82123
Juvenal - Satires - 11	946	0.90474
Juvenal - Satires - 12	571	0.86562
Juvenal - Satires - 13	1162	0.92309
Juvenal - Satires - 14	1524	0.93337
Juvenal - Satires - 15	811	0.84064
Juvenal - Satires - 16	267	0.99379

Persius - Satires - Prologus	46	0.97901
Persius - Satires - 1	619	0.89848
Persius - Satires - 2	369	0.86852
Persius - Satires - 3	573	0.91072
Persius - Satires - 4	234	0.88717
Persius - Satires - 5	923	0.90958
Persius - Satires - 6	381	0.91222
Horace - Satires - 1.1	631	0.96898
Horace - Satires - 1.2	694	0.96183
Horace - Satires - 1.3	706	0.96771
Horace - Satires - 1.4	735	0.95786
Horace - Satires - 1.5	499	0.97593
Horace - Satires - 1.6	679	0.90321
Horace - Satires - 1.7	164	0.99656
Horace - Satires - 1.8	239	0.98499
Horace - Satires - 1.9	414	0.90720
Horace - Satires - 1.10	477	0.95437
Horace - Satires - 2.1	421	0.95588
Horace - Satires - 2.2	694	0.96389
Horace - Satires - 2.3	1657	0.76242
Horace - Satires - 2.4	448	0.98510
Horace - Satires - 2.5	568	0.92755
Horace - Satires - 2.6	612	0.95161
Horace - Satires - 2.7	592	0.86800
Horace - Satires - 2.8	461	0.96239

Book 6 252 - 289		
Poems	**Poem Length**	**Coefficient**
Lucilius - Satires	149	1.00000
Juvenal - Satires - 1	770	0.98095
Juvenal - Satires - 2	737	0.98484
Juvenal - Satires - 3	1419	0.89164
Juvenal - Satires - 4	671	0.97520
Juvenal - Satires - 5	752	0.92152

Juvenal - Satires - 6	3085	0.91396
Juvenal - Satires - 7	1096	0.99364
Juvenal - Satires - 8	1169	0.95974
Juvenal - Satires - 9	696	0.92300
Juvenal - Satires - 10	1689	0.86862
Juvenal - Satires - 11	946	0.95459
Juvenal - Satires - 12	571	0.95836
Juvenal - Satires - 13	1162	0.96290
Juvenal - Satires - 14	1524	0.97161
Juvenal - Satires - 15	811	0.92798
Juvenal - Satires - 16	267	0.93666
Persius - Satires - Prologus	46	0.90256
Persius - Satires - 1	619	0.91157
Persius - Satires - 2	369	0.96882
Persius - Satires - 3	573	0.93477
Persius - Satires - 4	234	0.92652
Persius - Satires - 5	923	0.92347
Persius - Satires - 6	381	0.78447
Horace - Satires - 1.1	631	0.90609
Horace - Satires - 1.2	694	0.95412
Horace - Satires - 1.3	706	0.88098
Horace - Satires - 1.4	735	0.88230
Horace - Satires - 1.5	499	0.94089
Horace - Satires - 1.6	679	0.73563
Horace - Satires - 1.7	164	0.89328
Horace - Satires - 1.8	239	0.96556
Horace - Satires - 1.9	414	0.80297
Horace - Satires - 1.10	477	0.90556
Horace - Satires - 2.1	421	0.87883
Horace - Satires - 2.2	694	0.90901
Horace - Satires - 2.3	1657	0.80529
Horace - Satires - 2.4	448	0.90082
Horace - Satires - 2.5	568	0.83736
Horace - Satires - 2.6	612	0.89076
Horace - Satires - 2.7	592	0.80507
Horace - Satires - 2.8	461	0.88082

Book 7 290 - 323		
Poems	**Poem Length**	**Coefficient**
Lucilius - Satires	127	1.00000
Juvenal - Satires - 1	770	0.88945
Juvenal - Satires - 2	737	0.92117
Juvenal - Satires - 3	1419	0.88039
Juvenal - Satires - 4	671	0.92034
Juvenal - Satires - 5	752	0.95767
Juvenal - Satires - 6	3085	0.80186
Juvenal - Satires - 7	1096	0.94862
Juvenal - Satires - 8	1169	0.96636
Juvenal - Satires - 9	696	0.97430
Juvenal - Satires - 10	1689	0.77268
Juvenal - Satires - 11	946	0.88364
Juvenal - Satires - 12	571	0.85730
Juvenal - Satires - 13	1162	0.89519
Juvenal - Satires - 14	1524	0.91672
Juvenal - Satires - 15	811	0.81197
Juvenal - Satires - 16	267	0.95157
Persius - Satires - Prologus	46	0.90868
Persius - Satires - 1	619	0.97706
Persius - Satires - 2	369	0.96509
Persius - Satires - 3	573	0.98509
Persius - Satires - 4	234	0.97650
Persius - Satires - 5	923	0.98577
Persius - Satires - 6	381	0.92787
Horace - Satires - 1.1	631	0.98668
Horace - Satires - 1.2	694	0.99982
Horace - Satires - 1.3	706	0.97778
Horace - Satires - 1.4	735	0.97969
Horace - Satires - 1.5	499	0.96316
Horace - Satires - 1.6	679	0.88966
Horace - Satires - 1.7	164	0.95744

Horace - Satires - 1.8	239	0.96798
Horace - Satires - 1.9	414	0.93868
Horace - Satires - 1.10	477	0.98711
Horace - Satires - 2.1	421	0.97722
Horace - Satires - 2.2	694	0.98756
Horace - Satires - 2.3	1657	0.87930
Horace - Satires - 2.4	448	0.97231
Horace - Satires - 2.5	568	0.95894
Horace - Satires - 2.6	612	0.98355
Horace - Satires - 2.7	592	0.92772
Horace - Satires - 2.8	461	0.97680

Book 8 324 - 346		
Poems	**Poem Length**	**Coefficient**
Lucilius - Satires	78	1.00000
Juvenal - Satires - 1	770	0.92615
Juvenal - Satires - 2	737	0.97749
Juvenal - Satires - 3	1419	0.93124
Juvenal - Satires - 4	671	0.93136
Juvenal - Satires - 5	752	0.98257
Juvenal - Satires - 6	3085	0.91075
Juvenal - Satires - 7	1096	0.95861
Juvenal - Satires - 8	1169	0.98576
Juvenal - Satires - 9	696	0.97944
Juvenal - Satires - 10	1689	0.74056
Juvenal - Satires - 11	946	0.94009
Juvenal - Satires - 12	571	0.91431
Juvenal - Satires - 13	1162	0.91002
Juvenal - Satires - 14	1524	0.94961
Juvenal - Satires - 15	811	0.84632
Juvenal - Satires - 16	267	0.92339
Persius - Satires - Prologus	46	0.89636
Persius - Satires - 1	619	0.95599
Persius - Satires - 2	369	0.95855

Persius - Satires - 3	573	0.96850
Persius - Satires - 4	234	0.95917
Persius - Satires - 5	923	0.93709
Persius - Satires - 6	381	0.85024
Horace - Satires - 1.1	631	0.91113
Horace - Satires - 1.2	694	0.96431
Horace - Satires - 1.3	706	0.90969
Horace - Satires - 1.4	735	0.91819
Horace - Satires - 1.5	499	0.97011
Horace - Satires - 1.6	679	0.81755
Horace - Satires - 1.7	164	0.88753
Horace - Satires - 1.8	239	0.93632
Horace - Satires - 1.9	414	0.86790
Horace - Satires - 1.10	477	0.94485
Horace - Satires - 2.1	421	0.89246
Horace - Satires - 2.2	694	0.94609
Horace - Satires - 2.3	1657	0.76851
Horace - Satires - 2.4	448	0.93673
Horace - Satires - 2.5	568	0.88690
Horace - Satires - 2.6	612	0.92589
Horace - Satires - 2.7	592	0.89268
Horace - Satires - 2.8	461	0.89084

Book 9 347 - 410		
Poems	**Poem Length**	**Coefficient**
Lucilius - Satires	287	1.00000
Juvenal - Satires - 1	770	0.84842
Juvenal - Satires - 2	737	0.86147
Juvenal - Satires - 3	1419	0.73710
Juvenal - Satires - 4	671	0.85968
Juvenal - Satires - 5	752	0.86819
Juvenal - Satires - 6	3085	0.78520
Juvenal - Satires - 7	1096	0.90479
Juvenal - Satires - 8	1169	0.88293

Juvenal - Satires - 9	696	0.90784
Juvenal - Satires - 10	1689	0.75037
Juvenal - Satires - 11	946	0.78635
Juvenal - Satires - 12	571	0.78779
Juvenal - Satires - 13	1162	0.82895
Juvenal - Satires - 14	1524	0.83449
Juvenal - Satires - 15	811	0.75705
Juvenal - Satires - 16	267	0.87386
Persius - Satires - Prologus	46	0.81349
Persius - Satires - 1	619	0.95793
Persius - Satires - 2	369	0.97712
Persius - Satires - 3	573	0.96466
Persius - Satires - 4	234	0.96962
Persius - Satires - 5	923	0.98800
Persius - Satires - 6	381	0.87209
Horace - Satires - 1.1	631	0.95991
Horace - Satires - 1.2	694	0.97034
Horace - Satires - 1.3	706	0.93550
Horace - Satires - 1.4	735	0.94423
Horace - Satires - 1.5	499	0.87562
Horace - Satires - 1.6	679	0.81256
Horace - Satires - 1.7	164	0.89482
Horace - Satires - 1.8	239	0.91447
Horace - Satires - 1.9	414	0.89532
Horace - Satires - 1.10	477	0.95063
Horace - Satires - 2.1	421	0.95500
Horace - Satires - 2.2	694	0.93369
Horace - Satires - 2.3	1657	0.96605
Horace - Satires - 2.4	448	0.89593
Horace - Satires - 2.5	568	0.92207
Horace - Satires - 2.6	612	0.94648
Horace - Satires - 2.7	592	0.87852
Horace - Satires - 2.8	461	0.95272

Book 10 411 - 423		
Poems	**Poem Length**	**Coefficient**
Lucilius - Satires	43	1.00000
Juvenal - Satires - 1	770	0.90144
Juvenal - Satires - 2	737	0.86825
Juvenal - Satires - 3	1419	0.82814
Juvenal - Satires - 4	671	0.94444
Juvenal - Satires - 5	752	0.83711
Juvenal - Satires - 6	3085	0.69824
Juvenal - Satires - 7	1096	0.94425
Juvenal - Satires - 8	1169	0.89725
Juvenal - Satires - 9	696	0.84996
Juvenal - Satires - 10	1689	0.91842
Juvenal - Satires - 11	946	0.87320
Juvenal - Satires - 12	571	0.87201
Juvenal - Satires - 13	1162	0.93586
Juvenal - Satires - 14	1524	0.91354
Juvenal - Satires - 15	811	0.88491
Juvenal - Satires - 16	267	0.96143
Persius - Satires - Prologus	46	0.92823
Persius - Satires - 1	619	0.84580
Persius - Satires - 2	369	0.88167
Persius - Satires - 3	573	0.86739
Persius - Satires - 4	234	0.85095
Persius - Satires - 5	923	0.89302
Persius - Satires - 6	381	0.82175
Horace - Satires - 1.1	631	0.94348
Horace - Satires - 1.2	694	0.93225
Horace - Satires - 1.3	706	0.91272
Horace - Satires - 1.4	735	0.89852
Horace - Satires - 1.5	499	0.90305
Horace - Satires - 1.6	679	0.78541
Horace - Satires - 1.7	164	0.96227
Horace - Satires - 1.8	239	0.97904

Horace - Satires - 1.9	414	0.82206
Horace - Satires - 1.10	477	0.89265
Horace - Satires - 2.1	421	0.92360
Horace - Satires - 2.2	694	0.90105
Horace - Satires - 2.3	1657	0.82754
Horace - Satires - 2.4	448	0.91003
Horace - Satires - 2.5	568	0.85906
Horace - Satires - 2.6	612	0.89598
Horace - Satires - 2.7	592	0.76474
Horace - Satires - 2.8	461	0.93184

Book 11 424 - 454		
Poems	**Poem Length**	**Coefficient**
Lucilius - Satires	113	1.00000
Juvenal - Satires - 1	770	0.92707
Juvenal - Satires - 2	737	0.93921
Juvenal - Satires - 3	1419	0.94489
Juvenal - Satires - 4	671	0.97287
Juvenal - Satires - 5	752	0.94474
Juvenal - Satires - 6	3085	0.78100
Juvenal - Satires - 7	1096	0.97437
Juvenal - Satires - 8	1169	0.98025
Juvenal - Satires - 9	696	0.94045
Juvenal - Satires - 10	1689	0.85798
Juvenal - Satires - 11	946	0.94745
Juvenal - Satires - 12	571	0.92055
Juvenal - Satires - 13	1162	0.96031
Juvenal - Satires - 14	1524	0.97008
Juvenal - Satires - 15	811	0.89439
Juvenal - Satires - 16	267	0.99702
Persius - Satires - Prologus	46	0.98130
Persius - Satires - 1	619	0.90556
Persius - Satires - 2	369	0.90476
Persius - Satires - 3	573	0.92275

Persius - Satires - 4	234	0.90191
Persius - Satires - 5	923	0.91524
Persius - Satires - 6	381	0.87577
Horace - Satires - 1.1	631	0.95617
Horace - Satires - 1.2	694	0.96778
Horace - Satires - 1.3	706	0.94817
Horace - Satires - 1.4	735	0.94070
Horace - Satires - 1.5	499	0.98484
Horace - Satires - 1.6	679	0.85811
Horace - Satires - 1.7	164	0.97971
Horace - Satires - 1.8	239	0.99681
Horace - Satires - 1.9	414	0.87666
Horace - Satires - 1.10	477	0.94646
Horace - Satires - 2.1	421	0.93629
Horace - Satires - 2.2	694	0.95553
Horace - Satires - 2.3	1657	0.76332
Horace - Satires - 2.4	448	0.97336
Horace - Satires - 2.5	568	0.90255
Horace - Satires - 2.6	612	0.93819
Horace - Satires - 2.7	592	0.84988
Horace - Satires - 2.8	461	0.94231

Book 12 455 - 464		
Poems	**Poem Length**	**Coefficient**
Lucilius - Satires	38	1.00000
Juvenal - Satires - 1	770	0.88902
Juvenal - Satires - 2	737	0.92122
Juvenal - Satires - 3	1419	0.87667
Juvenal - Satires - 4	671	0.91814
Juvenal - Satires - 5	752	0.95697
Juvenal - Satires - 6	3085	0.80510
Juvenal - Satires - 7	1096	0.94765
Juvenal - Satires - 8	1169	0.96486
Juvenal - Satires - 9	696	0.97435

Juvenal - Satires - 10	1689	0.76952
Juvenal - Satires - 11	946	0.88137
Juvenal - Satires - 12	571	0.85576
Juvenal - Satires - 13	1162	0.89260
Juvenal - Satires - 14	1524	0.91453
Juvenal - Satires - 15	811	0.80958
Juvenal - Satires - 16	267	0.94808
Persius - Satires - Prologus	46	0.90396
Persius - Satires - 1	619	0.97875
Persius - Satires - 2	369	0.96777
Persius - Satires - 3	573	0.98664
Persius - Satires - 4	234	0.97872
Persius - Satires - 5	923	0.98741
Persius - Satires - 6	381	0.92666
Horace - Satires - 1.1	631	0.98576
Horace - Satires - 1.2	694	0.99979
Horace - Satires - 1.3	706	0.97653
Horace - Satires - 1.4	735	0.97887
Horace - Satires - 1.5	499	0.96087
Horace - Satires - 1.6	679	0.88720
Horace - Satires - 1.7	164	0.95406
Horace - Satires - 1.8	239	0.96564
Horace - Satires - 1.9	414	0.93821
Horace - Satires - 1.10	477	0.98678
Horace - Satires - 2.1	421	0.97644
Horace - Satires - 2.2	694	0.98680
Horace - Satires - 2.3	1657	0.88304
Horace - Satires - 2.4	448	0.96999
Horace - Satires - 2.5	568	0.95852
Horace - Satires - 2.6	612	0.98312
Horace - Satires - 2.7	592	0.92848
Horace - Satires - 2.8	461	0.97576

Book 13 465 - 478		
Poems	**Poem Length**	**Coefficient**
Lucilius - Satires	58	1.00000
Juvenal - Satires - 1	770	0.90626
Juvenal - Satires - 2	737	0.86775
Juvenal - Satires - 3	1419	0.82858
Juvenal - Satires - 4	671	0.94877
Juvenal - Satires - 5	752	0.82584
Juvenal - Satires - 6	3085	0.69601
Juvenal - Satires - 7	1096	0.94460
Juvenal - Satires - 8	1169	0.89144
Juvenal - Satires - 9	696	0.83571
Juvenal - Satires - 10	1689	0.93237
Juvenal - Satires - 11	946	0.87839
Juvenal - Satires - 12	571	0.87979
Juvenal - Satires - 13	1162	0.94255
Juvenal - Satires - 14	1524	0.91689
Juvenal - Satires - 15	811	0.89721
Juvenal - Satires - 16	267	0.96054
Persius - Satires - Prologus	46	0.93063
Persius - Satires - 1	619	0.82745
Persius - Satires - 2	369	0.86941
Persius - Satires - 3	573	0.85119
Persius - Satires - 4	234	0.83322
Persius - Satires - 5	923	0.87611
Persius - Satires - 6	381	0.80148
Horace - Satires - 1.1	631	0.93052
Horace - Satires - 1.2	694	0.92020
Horace - Satires - 1.3	706	0.89810
Horace - Satires - 1.4	735	0.88270
Horace - Satires - 1.5	499	0.89745
Horace - Satires - 1.6	679	0.76665
Horace - Satires - 1.7	164	0.95649
Horace - Satires - 1.8	239	0.97730

Horace - Satires - 1.9	414	0.80092
Horace - Satires - 1.10	477	0.87714
Horace - Satires - 2.1	421	0.90843
Horace - Satires - 2.2	694	0.88662
Horace - Satires - 2.3	1657	0.80778
Horace - Satires - 2.4	448	0.89985
Horace - Satires - 2.5	568	0.83985
Horace - Satires - 2.6	612	0.87954
Horace - Satires - 2.7	592	0.74184
Horace - Satires - 2.8	461	0.91754

Book 14 479 - 506		
Poems	**Poem Length**	**Coefficient**
Lucilius - Satires	115	1.00000
Juvenal - Satires - 1	770	0.77928
Juvenal - Satires - 2	737	0.83684
Juvenal - Satires - 3	1419	0.84604
Juvenal - Satires - 4	671	0.83548
Juvenal - Satires - 5	752	0.94851
Juvenal - Satires - 6	3085	0.69091
Juvenal - Satires - 7	1096	0.86813
Juvenal - Satires - 8	1169	0.93155
Juvenal - Satires - 9	696	0.97090
Juvenal - Satires - 10	1689	0.65119
Juvenal - Satires - 11	946	0.79838
Juvenal - Satires - 12	571	0.74779
Juvenal - Satires - 13	1162	0.80177
Juvenal - Satires - 14	1524	0.83696
Juvenal - Satires - 15	811	0.68677
Juvenal - Satires - 16	267	0.91579
Persius - Satires - Prologus	46	0.87286
Persius - Satires - 1	619	0.97653
Persius - Satires - 2	369	0.90875
Persius - Satires - 3	573	0.97273

Persius - Satires - 4	234	0.96314
Persius - Satires - 5	923	0.97529
Persius - Satires - 6	381	0.98499
Horace - Satires - 1.1	631	0.98857
Horace - Satires - 1.2	694	0.98047
Horace - Satires - 1.3	706	0.99646
Horace - Satires - 1.4	735	0.99942
Horace - Satires - 1.5	499	0.93985
Horace - Satires - 1.6	679	0.96061
Horace - Satires - 1.7	164	0.95211
Horace - Satires - 1.8	239	0.91840
Horace - Satires - 1.9	414	0.98982
Horace - Satires - 1.10	477	0.99728
Horace - Satires - 2.1	421	0.99239
Horace - Satires - 2.2	694	0.99598
Horace - Satires - 2.3	1657	0.86488
Horace - Satires - 2.4	448	0.97939
Horace - Satires - 2.5	568	0.99686
Horace - Satires - 2.6	612	0.99958
Horace - Satires - 2.7	592	0.97566
Horace - Satires - 2.8	461	0.98992

Book 15 507 - 543		
Poems	**Poem Length**	**Coefficient**
Lucilius - Satires	156	1.00000
Juvenal - Satires - 1	770	0.97520
Juvenal - Satires - 2	737	0.96959
Juvenal - Satires - 3	1419	0.86753
Juvenal - Satires - 4	671	0.97165
Juvenal - Satires - 5	752	0.90387
Juvenal - Satires - 6	3085	0.89079
Juvenal - Satires - 7	1096	0.99058
Juvenal - Satires - 8	1169	0.94682
Juvenal - Satires - 9	696	0.91138

Juvenal - Satires - 10	1689	0.88450
Juvenal - Satires - 11	946	0.93761
Juvenal - Satires - 12	571	0.94562
Juvenal - Satires - 13	1162	0.95930
Juvenal - Satires - 14	1524	0.96079
Juvenal - Satires - 15	811	0.92403
Juvenal - Satires - 16	267	0.93655
Persius - Satires - Prologus	46	0.89682
Persius - Satires - 1	619	0.90916
Persius - Satires - 2	369	0.97082
Persius - Satires - 3	573	0.93227
Persius - Satires - 4	234	0.92518
Persius - Satires - 5	923	0.93067
Persius - Satires - 6	381	0.78878
Horace - Satires - 1.1	631	0.91710
Horace - Satires - 1.2	694	0.95670
Horace - Satires - 1.3	706	0.88692
Horace - Satires - 1.4	735	0.88682
Horace - Satires - 1.5	499	0.92932
Horace - Satires - 1.6	679	0.73540
Horace - Satires - 1.7	164	0.90166
Horace - Satires - 1.8	239	0.96948
Horace - Satires - 1.9	414	0.80644
Horace - Satires - 1.10	477	0.90568
Horace - Satires - 2.1	421	0.89098
Horace - Satires - 2.2	694	0.90908
Horace - Satires - 2.3	1657	0.83529
Horace - Satires - 2.4	448	0.89774
Horace - Satires - 2.5	568	0.84278
Horace - Satires - 2.6	612	0.89500
Horace - Satires - 2.7	592	0.79922
Horace - Satires - 2.8	461	0.89359

Book 16 544 - 563		
Poems	**Poem Length**	**Coefficient**
Lucilius - Satires	79	1.00000
Juvenal - Satires - 1	770	0.90259
Juvenal - Satires - 2	737	0.90246
Juvenal - Satires - 3	1419	0.73909
Juvenal - Satires - 4	671	0.87948
Juvenal - Satires - 5	752	0.85012
Juvenal - Satires - 6	3085	0.86827
Juvenal - Satires - 7	1096	0.92350
Juvenal - Satires - 8	1169	0.87369
Juvenal - Satires - 9	696	0.88126
Juvenal - Satires - 10	1689	0.78989
Juvenal - Satires - 11	946	0.82468
Juvenal - Satires - 12	571	0.84318
Juvenal - Satires - 13	1162	0.85707
Juvenal - Satires - 14	1524	0.86046
Juvenal - Satires - 15	811	0.81479
Juvenal - Satires - 16	267	0.84312
Persius - Satires - Prologus	46	0.77376
Persius - Satires - 1	619	0.92451
Persius - Satires - 2	369	0.98967
Persius - Satires - 3	573	0.94019
Persius - Satires - 4	234	0.94735
Persius - Satires - 5	923	0.95443
Persius - Satires - 6	381	0.78211
Horace - Satires - 1.1	631	0.90119
Horace - Satires - 1.2	694	0.93882
Horace - Satires - 1.3	706	0.86372
Horace - Satires - 1.4	735	0.87153
Horace - Satires - 1.5	499	0.84754
Horace - Satires - 1.6	679	0.70368
Horace - Satires - 1.7	164	0.83291
Horace - Satires - 1.8	239	0.89796

Horace - Satires - 1.9	414	0.80989
Horace - Satires - 1.10	477	0.89023
Horace - Satires - 2.1	421	0.88500
Horace - Satires - 2.2	694	0.88408
Horace - Satires - 2.3	1657	0.93033
Horace - Satires - 2.4	448	0.83792
Horace - Satires - 2.5	568	0.84375
Horace - Satires - 2.6	612	0.88603
Horace - Satires - 2.7	592	0.81209
Horace - Satires - 2.8	461	0.88166

Book 17 564 - 580		
Poems	**Poem Length**	**Coefficient**
Lucilius - Satires	62	1.00000
Juvenal - Satires - 1	770	0.89958
Juvenal - Satires - 2	737	0.94355
Juvenal - Satires - 3	1419	0.97640
Juvenal - Satires - 4	671	0.94627
Juvenal - Satires - 5	752	0.98211
Juvenal - Satires - 6	3085	0.79909
Juvenal - Satires - 7	1096	0.95218
Juvenal - Satires - 8	1169	0.99604
Juvenal - Satires - 9	696	0.97047
Juvenal - Satires - 10	1689	0.77323
Juvenal - Satires - 11	946	0.94845
Juvenal - Satires - 12	571	0.90421
Juvenal - Satires - 13	1162	0.92965
Juvenal - Satires - 14	1524	0.96126
Juvenal - Satires - 15	811	0.85139
Juvenal - Satires - 16	267	0.97980
Persius - Satires - Prologus	46	0.97210
Persius - Satires - 1	619	0.92210
Persius - Satires - 2	369	0.89207
Persius - Satires - 3	573	0.93430

Persius - Satires - 4	234	0.91216
Persius - Satires - 5	923	0.90752
Persius - Satires - 6	381	0.89302
Horace - Satires - 1.1	631	0.93806
Horace - Satires - 1.2	694	0.96130
Horace - Satires - 1.3	706	0.94573
Horace - Satires - 1.4	735	0.94349
Horace - Satires - 1.5	499	0.99993
Horace - Satires - 1.6	679	0.88657
Horace - Satires - 1.7	164	0.95741
Horace - Satires - 1.8	239	0.96887
Horace - Satires - 1.9	414	0.89491
Horace - Satires - 1.10	477	0.95541
Horace - Satires - 2.1	421	0.92093
Horace - Satires - 2.2	694	0.96378
Horace - Satires - 2.3	1657	0.71234
Horace - Satires - 2.4	448	0.98274
Horace - Satires - 2.5	568	0.91119
Horace - Satires - 2.6	612	0.94147
Horace - Satires - 2.7	592	0.88934
Horace - Satires - 2.8	461	0.92464

Book 18 581 - 583		
Poems	**Poem Length**	**Coefficient**
Lucilius - Satires	8	1.00000
Juvenal - Satires - 1	770	0.31553
Juvenal - Satires - 2	737	0.28377
Juvenal - Satires - 3	1419	0.03923
Juvenal - Satires - 4	671	0.27990
Juvenal - Satires - 5	752	0.28466
Juvenal - Satires - 6	3085	0.30947
Juvenal - Satires - 7	1096	0.35737
Juvenal - Satires - 8	1169	0.27697
Juvenal - Satires - 9	696	0.37848

Juvenal - Satires - 10	1689	0.31269
Juvenal - Satires - 11	946	0.14158
Juvenal - Satires - 12	571	0.18923
Juvenal - Satires - 13	1162	0.24540
Juvenal - Satires - 14	1524	0.21796
Juvenal - Satires - 15	811	0.20004
Juvenal - Satires - 16	267	0.28356
Persius - Satires - Prologus	46	0.15875
Persius - Satires - 1	619	0.53737
Persius - Satires - 2	369	0.59280
Persius - Satires - 3	573	0.52933
Persius - Satires - 4	234	0.56690
Persius - Satires - 5	923	0.61673
Persius - Satires - 6	381	0.45857
Horace - Satires - 1.1	631	0.53050
Horace - Satires - 1.2	694	0.50181
Horace - Satires - 1.3	706	0.47375
Horace - Satires - 1.4	735	0.48547
Horace - Satires - 1.5	499	0.25463
Horace - Satires - 1.6	679	0.34570
Horace - Satires - 1.7	164	0.37327
Horace - Satires - 1.8	239	0.37462
Horace - Satires - 1.9	414	0.48957
Horace - Satires - 1.10	477	0.47299
Horace - Satires - 2.1	421	0.54937
Horace - Satires - 2.2	694	0.44942
Horace - Satires - 2.3	1657	0.85498
Horace - Satires - 2.4	448	0.33996
Horace - Satires - 2.5	568	0.51138
Horace - Satires - 2.6	612	0.50437
Horace - Satires - 2.7	592	0.46387
Horace - Satires - 2.8	461	0.53746

Book 19 584 - 594		
Poems	**Poem Length**	**Coefficient**
Lucilius - Satires	58	1.00000
Juvenal - Satires - 1	770	0.86253
Juvenal - Satires - 2	737	0.87832
Juvenal - Satires - 3	1419	0.84763
Juvenal - Satires - 4	671	0.90741
Juvenal - Satires - 5	752	0.92031
Juvenal - Satires - 6	3085	0.72831
Juvenal - Satires - 7	1096	0.93019
Juvenal - Satires - 8	1169	0.93806
Juvenal - Satires - 9	696	0.94237
Juvenal - Satires - 10	1689	0.79356
Juvenal - Satires - 11	946	0.85038
Juvenal - Satires - 12	571	0.82544
Juvenal - Satires - 13	1162	0.88367
Juvenal - Satires - 14	1524	0.89245
Juvenal - Satires - 15	811	0.79717
Juvenal - Satires - 16	267	0.95429
Persius - Satires - Prologus	46	0.90950
Persius - Satires - 1	619	0.95111
Persius - Satires - 2	369	0.93670
Persius - Satires - 3	573	0.95907
Persius - Satires - 4	234	0.94852
Persius - Satires - 5	923	0.97456
Persius - Satires - 6	381	0.93299
Horace - Satires - 1.1	631	0.99642
Horace - Satires - 1.2	694	0.98935
Horace - Satires - 1.3	706	0.98329
Horace - Satires - 1.4	735	0.97980
Horace - Satires - 1.5	499	0.94227
Horace - Satires - 1.6	679	0.89621
Horace - Satires - 1.7	164	0.97559
Horace - Satires - 1.8	239	0.96950

Horace - Satires - 1.9	414	0.93850
Horace - Satires - 1.10	477	0.97727
Horace - Satires - 2.1	421	0.98927
Horace - Satires - 2.2	694	0.97903
Horace - Satires - 2.3	1657	0.89400
Horace - Satires - 2.4	448	0.96763
Horace - Satires - 2.5	568	0.96069
Horace - Satires - 2.6	612	0.98046
Horace - Satires - 2.7	592	0.90648
Horace - Satires - 2.8	461	0.99087

Book 20 595 - 622		
Poems	**Poem Length**	**Coefficient**
Lucilius - Satires	92	1.00000
Juvenal - Satires - 1	770	0.96347
Juvenal - Satires - 2	737	0.95230
Juvenal - Satires - 3	1419	0.92391
Juvenal - Satires - 4	671	0.99323
Juvenal - Satires - 5	752	0.90672
Juvenal - Satires - 6	3085	0.81073
Juvenal - Satires - 7	1096	0.99026
Juvenal - Satires - 8	1169	0.96097
Juvenal - Satires - 9	696	0.90050
Juvenal - Satires - 10	1689	0.91779
Juvenal - Satires - 11	946	0.96262
Juvenal - Satires - 12	571	0.95483
Juvenal - Satires - 13	1162	0.98702
Juvenal - Satires - 14	1524	0.98320
Juvenal - Satires - 15	811	0.94488
Juvenal - Satires - 16	267	0.98643
Persius - Satires - Prologus	46	0.96961
Persius - Satires - 1	619	0.86734
Persius - Satires - 2	369	0.90495
Persius - Satires - 3	573	0.89241

Persius - Satires - 4	234	0.87233
Persius - Satires - 5	923	0.88814
Persius - Satires - 6	381	0.80606
Horace - Satires - 1.1	631	0.92385
Horace - Satires - 1.2	694	0.94419
Horace - Satires - 1.3	706	0.90297
Horace - Satires - 1.4	735	0.89352
Horace - Satires - 1.5	499	0.95927
Horace - Satires - 1.6	679	0.77900
Horace - Satires - 1.7	164	0.95305
Horace - Satires - 1.8	239	0.99598
Horace - Satires - 1.9	414	0.80998
Horace - Satires - 1.10	477	0.90295
Horace - Satires - 2.1	421	0.89626
Horace - Satires - 2.2	694	0.91328
Horace - Satires - 2.3	1657	0.75303
Horace - Satires - 2.4	448	0.93181
Horace - Satires - 2.5	568	0.84478
Horace - Satires - 2.6	612	0.89299
Horace - Satires - 2.7	592	0.78105
Horace - Satires - 2.8	461	0.90391

Book 21 No fragments

Book 22 623 - 628		
Poems	**Poem Length**	**Coefficient**
Lucilius - Satires	19	1.00000
Juvenal - Satires - 1	770	0.90362
Juvenal - Satires - 2	737	0.96044
Juvenal - Satires - 3	1419	0.94373
Juvenal - Satires - 4	671	0.92559
Juvenal - Satires - 5	752	0.99442
Juvenal - Satires - 6	3085	0.86671
Juvenal - Satires - 7	1096	0.95022

Juvenal - Satires - 8	1169	0.99324
Juvenal - Satires - 9	696	0.99199
Juvenal - Satires - 10	1689	0.72736
Juvenal - Satires - 11	946	0.93099
Juvenal - Satires - 12	571	0.89415
Juvenal - Satires - 13	1162	0.90252
Juvenal - Satires - 14	1524	0.94403
Juvenal - Satires - 15	811	0.82517
Juvenal - Satires - 16	267	0.94130
Persius - Satires - Prologus	46	0.91718
Persius - Satires - 1	619	0.96496
Persius - Satires - 2	369	0.94508
Persius - Satires - 3	573	0.97411
Persius - Satires - 4	234	0.96180
Persius - Satires - 5	923	0.94520
Persius - Satires - 6	381	0.89184
Horace - Satires - 1.1	631	0.93522
Horace - Satires - 1.2	694	0.97536
Horace - Satires - 1.3	706	0.93985
Horace - Satires - 1.4	735	0.94623
Horace - Satires - 1.5	499	0.98509
Horace - Satires - 1.6	679	0.86768
Horace - Satires - 1.7	164	0.91789
Horace - Satires - 1.8	239	0.94588
Horace - Satires - 1.9	414	0.90452
Horace - Satires - 1.10	477	0.96678
Horace - Satires - 2.1	421	0.92052
Horace - Satires - 2.2	694	0.96904
Horace - Satires - 2.3	1657	0.77032
Horace - Satires - 2.4	448	0.96477
Horace - Satires - 2.5	568	0.91970
Horace - Satires - 2.6	612	0.95088
Horace - Satires - 2.7	592	0.92079
Horace - Satires - 2.8	461	0.91949

Book 23 629 -		
Poems	**Poem Length**	**Coefficient**
Lucilius - Satires	3	1.00000
Juvenal - Satires - 1	770	0.84102
Juvenal - Satires - 2	737	0.86271
Juvenal - Satires - 3	1419	0.92066
Juvenal - Satires - 4	671	0.91660
Juvenal - Satires - 5	752	0.92594
Juvenal - Satires - 6	3085	0.66015
Juvenal - Satires - 7	1096	0.91492
Juvenal - Satires - 8	1169	0.95034
Juvenal - Satires - 9	696	0.92570
Juvenal - Satires - 10	1689	0.79329
Juvenal - Satires - 11	946	0.88260
Juvenal - Satires - 12	571	0.83614
Juvenal - Satires - 13	1162	0.90032
Juvenal - Satires - 14	1524	0.91279
Juvenal - Satires - 15	811	0.80949
Juvenal - Satires - 16	267	0.98685
Persius - Satires - Prologus	46	0.97466
Persius - Satires - 1	619	0.88945
Persius - Satires - 2	369	0.84501
Persius - Satires - 3	573	0.89906
Persius - Satires - 4	234	0.87390
Persius - Satires - 5	923	0.89882
Persius - Satires - 6	381	0.92396
Horace - Satires - 1.1	631	0.96681
Horace - Satires - 1.2	694	0.95113
Horace - Satires - 1.3	706	0.96989
Horace - Satires - 1.4	735	0.95890
Horace - Satires - 1.5	499	0.96896
Horace - Satires - 1.6	679	0.92140
Horace - Satires - 1.7	164	0.99756
Horace - Satires - 1.8	239	0.97205

Horace - Satires - 1.9	414	0.91577
Horace - Satires - 1.10	477	0.95189
Horace - Satires - 2.1	421	0.95632
Horace - Satires - 2.2	694	0.96203
Horace - Satires - 2.3	1657	0.74810
Horace - Satires - 2.4	448	0.98624
Horace - Satires - 2.5	568	0.93304
Horace - Satires - 2.6	612	0.95096
Horace - Satires - 2.7	592	0.87335
Horace - Satires - 2.8	461	0.96293

Book 24 No fragments

Book 25 630 - 631		
Poems	**Poem Length**	**Coefficient**
Juvenal - Satires - 1	770	1.00000
Juvenal - Satires - 2	737	0.98016
Juvenal - Satires - 3	1419	0.88303
Juvenal - Satires - 4	671	0.98418
Juvenal - Satires - 5	752	0.85974
Juvenal - Satires - 6	3085	0.91641
Juvenal - Satires - 7	1096	0.98643
Juvenal - Satires - 8	1169	0.92243
Juvenal - Satires - 9	696	0.84700
Juvenal - Satires - 10	1689	0.91935
Juvenal - Satires - 11	946	0.97281
Juvenal - Satires - 12	571	0.99133
Juvenal - Satires - 13	1162	0.98349
Juvenal - Satires - 14	1524	0.97910
Juvenal - Satires - 15	811	0.97940
Juvenal - Satires - 16	267	0.91281
Persius - Satires - Prologus	46	0.89369
Persius - Satires - 1	619	0.81624

Persius - Satires - 2	369	0.90921
Persius - Satires - 3	573	0.84970
Persius - Satires - 4	234	0.83673
Persius - Satires - 5	923	0.83221
Persius - Satires - 6	381	0.66761
Horace - Satires - 1.1	631	0.82450
Horace - Satires - 1.2	694	0.88393
Horace - Satires - 1.3	706	0.79159
Horace - Satires - 1.4	735	0.78853
Horace - Satires - 1.5	499	0.90059
Horace - Satires - 1.6	679	0.62306
Horace - Satires - 1.7	164	0.84086
Horace - Satires - 1.8	239	0.94059
Horace - Satires - 1.9	414	0.68440
Horace - Satires - 1.10	477	0.81644
Horace - Satires - 2.1	421	0.78583
Horace - Satires - 2.2	694	0.82457
Horace - Satires - 2.3	1657	0.70041
Horace - Satires - 2.4	448	0.83289
Horace - Satires - 2.5	568	0.72719
Horace - Satires - 2.6	612	0.79599
Horace - Satires - 2.7	592	0.68383
Horace - Satires - 2.8	461	0.79174

Book 26 632 - 736		
Poems	**Poem Length**	**Coefficient**
Lucilius - Satires	498	1.00000
Juvenal - Satires - 1	770	0.80355
Juvenal - Satires - 2	737	0.86122
Juvenal - Satires - 3	1419	0.87189
Juvenal - Satires - 4	671	0.85663
Juvenal - Satires - 5	752	0.96277
Juvenal - Satires - 6	3085	0.72035
Juvenal - Satires - 7	1096	0.88927

Juvenal - Satires - 8	1169	0.94893
Juvenal - Satires - 9	696	0.98007
Juvenal - Satires - 10	1689	0.67696
Juvenal - Satires - 11	946	0.82669
Juvenal - Satires - 12	571	0.77813
Juvenal - Satires - 13	1162	0.82902
Juvenal - Satires - 14	1524	0.86182
Juvenal - Satires - 15	811	0.72220
Juvenal - Satires - 16	267	0.93164
Persius - Satires - Prologus	46	0.89257
Persius - Satires - 1	619	0.97894
Persius - Satires - 2	369	0.92163
Persius - Satires - 3	573	0.97771
Persius - Satires - 4	234	0.96666
Persius - Satires - 5	923	0.97532
Persius - Satires - 6	381	0.97785
Horace - Satires - 1.1	631	0.98893
Horace - Satires - 1.2	694	0.98730
Horace - Satires - 1.3	706	0.99623
Horace - Satires - 1.4	735	0.99883
Horace - Satires - 1.5	499	0.95469
Horace - Satires - 1.6	679	0.95293
Horace - Satires - 1.7	164	0.95888
Horace - Satires - 1.8	239	0.93349
Horace - Satires - 1.9	414	0.98278
Horace - Satires - 1.10	477	0.99947
Horace - Satires - 2.1	421	0.99008
Horace - Satires - 2.2	694	0.99929
Horace - Satires - 2.3	1657	0.85073
Horace - Satires - 2.4	448	0.98614
Horace - Satires - 2.5	568	0.99175
Horace - Satires - 2.6	612	0.99925
Horace - Satires - 2.7	592	0.97185
Horace - Satires - 2.8	461	0.98838

Book 27 737 - 792		
Poems	**Poem Length**	**Coefficient**
Lucilius - Satires	278	1.00000
Juvenal - Satires - 1	770	0.82303
Juvenal - Satires - 2	737	0.87186
Juvenal - Satires - 3	1419	0.90481
Juvenal - Satires - 4	671	0.88656
Juvenal - Satires - 5	752	0.96477
Juvenal - Satires - 6	3085	0.70919
Juvenal - Satires - 7	1096	0.90482
Juvenal - Satires - 8	1169	0.96172
Juvenal - Satires - 9	696	0.97422
Juvenal - Satires - 10	1689	0.71405
Juvenal - Satires - 11	946	0.85780
Juvenal - Satires - 12	571	0.80792
Juvenal - Satires - 13	1162	0.85901
Juvenal - Satires - 14	1524	0.88889
Juvenal - Satires - 15	811	0.75424
Juvenal - Satires - 16	267	0.95917
Persius - Satires - Prologus	46	0.93076
Persius - Satires - 1	619	0.95779
Persius - Satires - 2	369	0.89673
Persius - Satires - 3	573	0.95954
Persius - Satires - 4	234	0.94329
Persius - Satires - 5	923	0.95462
Persius - Satires - 6	381	0.96914
Horace - Satires - 1.1	631	0.98686
Horace - Satires - 1.2	694	0.98289
Horace - Satires - 1.3	706	0.99564
Horace - Satires - 1.4	735	0.99447
Horace - Satires - 1.5	499	0.97280
Horace - Satires - 1.6	679	0.95426
Horace - Satires - 1.7	164	0.97914
Horace - Satires - 1.8	239	0.95385

Horace - Satires - 1.9	414	0.96994
Horace - Satires - 1.10	477	0.99319
Horace - Satires - 2.1	421	0.98489
Horace - Satires - 2.2	694	0.99641
Horace - Satires - 2.3	1657	0.81183
Horace - Satires - 2.4	448	0.99738
Horace - Satires - 2.5	568	0.98051
Horace - Satires - 2.6	612	0.99153
Horace - Satires - 2.7	592	0.95033
Horace - Satires - 2.8	461	0.98550

Book 28 793 - 851		
Poems	**Poem Length**	**Coefficient**
Lucilius - Satires	246	1.00000
Juvenal - Satires - 1	770	0.88441
Juvenal - Satires - 2	737	0.90470
Juvenal - Satires - 3	1419	0.83870
Juvenal - Satires - 4	671	0.90851
Juvenal - Satires - 5	752	0.92915
Juvenal - Satires - 6	3085	0.79303
Juvenal - Satires - 7	1096	0.94070
Juvenal - Satires - 8	1169	0.94180
Juvenal - Satires - 9	696	0.95375
Juvenal - Satires - 10	1689	0.78185
Juvenal - Satires - 11	946	0.85796
Juvenal - Satires - 12	571	0.84138
Juvenal - Satires - 13	1162	0.88233
Juvenal - Satires - 14	1524	0.89649
Juvenal - Satires - 15	811	0.80383
Juvenal - Satires - 16	267	0.93358
Persius - Satires - Prologus	46	0.88085
Persius - Satires - 1	619	0.97133
Persius - Satires - 2	369	0.97259
Persius - Satires - 3	573	0.97982

Persius - Satires - 4	234	0.97495
Persius - Satires - 5	923	0.99050
Persius - Satires - 6	381	0.91404
Horace - Satires - 1.1	631	0.98499
Horace - Satires - 1.2	694	0.99476
Horace - Satires - 1.3	706	0.96840
Horace - Satires - 1.4	735	0.97024
Horace - Satires - 1.5	499	0.93670
Horace - Satires - 1.6	679	0.86576
Horace - Satires - 1.7	164	0.94537
Horace - Satires - 1.8	239	0.95915
Horace - Satires - 1.9	414	0.92724
Horace - Satires - 1.10	477	0.97593
Horace - Satires - 2.1	421	0.97624
Horace - Satires - 2.2	694	0.97457
Horace - Satires - 2.3	1657	0.91582
Horace - Satires - 2.4	448	0.95112
Horace - Satires - 2.5	568	0.95054
Horace - Satires - 2.6	612	0.97569
Horace - Satires - 2.7	592	0.91168
Horace - Satires - 2.8	461	0.97539

Book 29 852 - 973		
Poems	**Poem Length**	**Coefficient**
Lucilius - Satires	494	1.00000
Juvenal - Satires - 1	770	0.73332
Juvenal - Satires - 2	737	0.80205
Juvenal - Satires - 3	1419	0.82596
Juvenal - Satires - 4	671	0.79074
Juvenal - Satires - 5	752	0.94075
Juvenal - Satires - 6	3085	0.66134
Juvenal - Satires - 7	1096	0.83004
Juvenal - Satires - 8	1169	0.91252
Juvenal - Satires - 9	696	0.96459

Juvenal - Satires - 10	1689	0.59943
Juvenal - Satires - 11	946	0.76122
Juvenal - Satires - 12	571	0.70318
Juvenal - Satires - 13	1162	0.75998
Juvenal - Satires - 14	1524	0.79880
Juvenal - Satires - 15	811	0.64174
Juvenal - Satires - 16	267	0.89184
Persius - Satires - Prologus	46	0.84894
Persius - Satires - 1	619	0.97162
Persius - Satires - 2	369	0.88411
Persius - Satires - 3	573	0.96447
Persius - Satires - 4	234	0.95370
Persius - Satires - 5	923	0.96395
Persius - Satires - 6	381	0.99261
Horace - Satires - 1.1	631	0.97651
Horace - Satires - 1.2	694	0.96479
Horace - Satires - 1.3	706	0.99072
Horace - Satires - 1.4	735	0.99529
Horace - Satires - 1.5	499	0.92144
Horace - Satires - 1.6	679	0.97152
Horace - Satires - 1.7	164	0.93652
Horace - Satires - 1.8	239	0.88445
Horace - Satires - 1.9	414	0.99673
Horace - Satires - 1.10	477	0.99116
Horace - Satires - 2.1	421	0.98519
Horace - Satires - 2.2	694	0.98840
Horace - Satires - 2.3	1657	0.85296
Horace - Satires - 2.4	448	0.96993
Horace - Satires - 2.5	568	0.99916
Horace - Satires - 2.6	612	0.99542
Horace - Satires - 2.7	592	0.98609
Horace - Satires - 2.8	461	0.98224

Book 30 1000 - 1130		
Poems	**Poem Length**	**Coefficient**
Lucilius - Satires	555	1.00000
Juvenal - Satires - 1	770	0.82030
Juvenal - Satires - 2	737	0.86227
Juvenal - Satires - 3	1419	0.79198
Juvenal - Satires - 4	671	0.84219
Juvenal - Satires - 5	752	0.92438
Juvenal - Satires - 6	3085	0.76020
Juvenal - Satires - 7	1096	0.89050
Juvenal - Satires - 8	1169	0.91455
Juvenal - Satires - 9	696	0.95582
Juvenal - Satires - 10	1689	0.69205
Juvenal - Satires - 11	946	0.79520
Juvenal - Satires - 12	571	0.76682
Juvenal - Satires - 13	1162	0.81185
Juvenal - Satires - 14	1524	0.83670
Juvenal - Satires - 15	811	0.71967
Juvenal - Satires - 16	267	0.88731
Persius - Satires - Prologus	46	0.82192
Persius - Satires - 1	619	0.98715
Persius - Satires - 2	369	0.96579
Persius - Satires - 3	573	0.98762
Persius - Satires - 4	234	0.98767
Persius - Satires - 5	923	0.99925
Persius - Satires - 6	381	0.93894
Horace - Satires - 1.1	631	0.98029
Horace - Satires - 1.2	694	0.98566
Horace - Satires - 1.3	706	0.96906
Horace - Satires - 1.4	735	0.97615
Horace - Satires - 1.5	499	0.90202
Horace - Satires - 1.6	679	0.88734
Horace - Satires - 1.7	164	0.91562
Horace - Satires - 1.8	239	0.91045

Horace - Satires - 1.9	414	0.95410
Horace - Satires - 1.10	477	0.98058
Horace - Satires - 2.1	421	0.97939
Horace - Satires - 2.2	694	0.97545
Horace - Satires - 2.3	1657	0.93852
Horace - Satires - 2.4	448	0.93790
Horace - Satires - 2.5	568	0.96988
Horace - Satires - 2.6	612	0.98340
Horace - Satires - 2.7	592	0.94674
Horace - Satires - 2.8	461	0.97492

Appendix E - Pearson coefficients of Moby Dick first paragraphs of random chapters.

Chapter 8: -0.06018
Chapter 13: -0.09362
Chapter 18: -0.10465
Chapter 28: -0.14638
Chapter 32: -0.12222
Chapter 35: 0.00532
Chapter 37: -0.08223
Chapter 53: -0.10309
Chapter 59: -0.13750
Chapter 60: -0.05667
Chapter 62: -0.08647
Chapter 63: -0.02885
Chapter 64: -0.09634
Chapter 65: 0.07172
Chapter 66: -0.02745
Chapter 68: -0.14105
Chapter 70: -0.01856
Chapter 72: 0.03298
Chapter 73: -0.00271
Chapter 74: 0.00863
Chapter 76: -0.10640
Chapter 77: -0.02177
Chapter 79: -0.13546
Chapter 81: -0.11826
Chapter 83: -0.05415
Chapter 86: -0.10268
Chapter 93: -0.16104
Chapter 105:: -0.03888
Chapter 108: -0.00160
Chapter 112: -0.19397

Appendix F - Unassigned fragments correlated against the books of Lucilius

Fragment ID 771		
Poems	**Poem Length**	**Coefficient**
Lucilius Fragment Group 771	10	1.00000
Lucilius - Book 1	188	0.83628
Lucilius - Book 2	129	0.78564
Lucilius - Book 3	185	0.92007
Lucilius - Book 4	155	0.89575
Lucilius - Book 5	246	0.96849
Lucilius - Book 6	149	0.80702
Lucilius - Book 7	127	0.89944
Lucilius - Book 8	78	0.94843
Lucilius - Book 9	287	0.70109
Lucilius - Book 10	43	0.80495
Lucilius - Book 11	113	0.93486
Lucilius - Book 12	38	0.74229
Lucilius - Book 13	58	0.88296
Lucilius - Book 14	115	0.93644
Lucilius - Book 15	156	0.85278
Lucilius - Book 16	79	0.70390
Lucilius - Book 17	62	0.94800
Lucilius - Book 18	8	0.61727
Lucilius - Book 19	58	0.91905
Lucilius - Book 20	92	0.77934
Lucilius - Book 22	19	0.75970
Lucilius - Book 23	3	0.19754
Lucilius - Book 25	0	0.00000
Lucilius - Book 26	498	0.53556

Lucilius - Book 27	278	0.98647
Lucilius - Book 28	246	0.91049
Lucilius - Book 29	494	0.90079
Lucilius - Book 30	555	0.55325

Fragment ID 772		
Poems	**Poem Length**	**Coefficient**
Lucilius Fragment Group 772	9	1.00000
Lucilius - Book 1	188	0.54920
Lucilius - Book 2	129	0.80984
Lucilius - Book 3	185	0.77430
Lucilius - Book 4	155	0.86933
Lucilius - Book 5	246	0.90060
Lucilius - Book 6	149	0.71864
Lucilius - Book 7	127	0.83600
Lucilius - Book 8	78	0.90764
Lucilius - Book 9	287	0.31448
Lucilius - Book 10	43	0.78653
Lucilius - Book 11	113	0.80431
Lucilius - Book 12	38	0.82855
Lucilius - Book 13	58	0.68416
Lucilius - Book 14	115	0.81746
Lucilius - Book 15	156	0.87528
Lucilius - Book 16	79	0.61500
Lucilius - Book 17	62	0.82248
Lucilius - Book 18	8	0.32083
Lucilius - Book 19	58	0.78939
Lucilius - Book 20	92	0.82377
Lucilius - Book 22	19	0.77061
Lucilius - Book 23	3	0.07161
Lucilius - Book 25	0	0.00000
Lucilius - Book 26	498	0.73234
Lucilius - Book 27	278	0.87787
Lucilius - Book 28	246	0.84468

Lucilius - Book 29	494	0.97737
Lucilius - Book 30	555	0.62810

Fragment ID 774		
Poems	**Poem Length**	**Coefficient**
Lucilius Fragment Group 774	44	1.00000
Lucilius - Book 1	188	0.84880
Lucilius - Book 2	129	0.94290
Lucilius - Book 3	185	0.97225
Lucilius - Book 4	155	0.96587
Lucilius - Book 5	246	0.98141
Lucilius - Book 6	149	0.92718
Lucilius - Book 7	127	0.98452
Lucilius - Book 8	78	0.99938
Lucilius - Book 9	287	0.69591
Lucilius - Book 10	43	0.94979
Lucilius - Book 11	113	0.98102
Lucilius - Book 12	38	0.92065
Lucilius - Book 13	58	0.93235
Lucilius - Book 14	115	0.98351
Lucilius - Book 15	156	0.97577
Lucilius - Book 16	79	0.83017
Lucilius - Book 17	62	0.98402
Lucilius - Book 18	8	0.51600
Lucilius - Book 19	58	0.97758
Lucilius - Book 20	92	0.94326
Lucilius - Book 22	19	0.74637
Lucilius - Book 23	3	0.40301
Lucilius - Book 25	0	0.00000
Lucilius - Book 26	498	0.75969
Lucilius - Book 27	278	0.97285
Lucilius - Book 28	246	0.96679
Lucilius - Book 29	494	0.91066

Lucilius - Book 30	555	0.73844

Fragment ID 811		
Poems	**Poem Length**	**Coefficient**
Lucilius Fragment Group 811	76	1.00000
Lucilius - Book 1	188	0.75254
Lucilius - Book 2	129	0.99463
Lucilius - Book 3	185	0.90894
Lucilius - Book 4	155	0.91650
Lucilius - Book 5	246	0.88385
Lucilius - Book 6	149	0.95007
Lucilius - Book 7	127	0.95322
Lucilius - Book 8	78	0.93927
Lucilius - Book 9	287	0.59866
Lucilius - Book 10	43	0.98140
Lucilius - Book 11	113	0.91303
Lucilius - Book 12	38	0.99166
Lucilius - Book 13	58	0.87937
Lucilius - Book 14	115	0.92552
Lucilius - Book 15	156	0.99362
Lucilius - Book 16	79	0.88207
Lucilius - Book 17	62	0.90570
Lucilius - Book 18	8	0.38958
Lucilius - Book 19	58	0.92982
Lucilius - Book 20	92	0.99481
Lucilius - Book 22	19	0.62942
Lucilius - Book 23	3	0.56355
Lucilius - Book 25	0	0.00000
Lucilius - Book 26	498	0.87933
Lucilius - Book 27	278	0.84895
Lucilius - Book 28	246	0.92778
Lucilius - Book 29	494	0.83131

Lucilius - Book 30	555	0.86179

Appendix G - Unassigned fragments suspected to be from books XXVI-XXIX

Unassigned Fragment Lines 981-999		
Poems	**Poem Length**	**Coefficient**
Lucilius Fragment Group Array	70	1.00000
Lucilius - Book 1	188	0.87971
Lucilius - Book 2	129	0.95682
Lucilius - Book 3	185	0.98543
Lucilius - Book 4	155	0.96236
Lucilius - Book 5	246	0.96885
Lucilius - Book 6	149	0.95253
Lucilius - Book 7	127	0.99205
Lucilius - Book 8	78	0.99273
Lucilius - Book 9	287	0.74452
Lucilius - Book 10	43	0.96702
Lucilius - Book 11	113	0.98965
Lucilius - Book 12	38	0.93165
Lucilius - Book 13	58	0.95595
Lucilius - Book 14	115	0.99106
Lucilius - Book 15	156	0.97994
Lucilius - Book 16	79	0.86116
Lucilius - Book 17	62	0.98852
Lucilius - Book 18	8	0.52835
Lucilius - Book 19	58	0.99008
Lucilius - Book 20	92	0.95372
Lucilius - Book 22	19	0.71683
Lucilius - Book 23	3	0.49653
Lucilius - Book 25	0	0.00000
Lucilius - Book 26	498	0.76630

Lucilius - Book 27	278	0.96000
Lucilius - Book 28	246	0.96795
Lucilius - Book 29	494	0.87362
Lucilius - Book 30	555	0.75931

Unassigned Fragment Lines 974-980		
Poems	**Poem Length**	**Coefficient**
Lucilius Fragment Group Array	23	1.00000
Lucilius - Book 1	188	0.88258
Lucilius - Book 2	129	0.87909
Lucilius - Book 3	185	0.95167
Lucilius - Book 4	155	0.85638
Lucilius - Book 5	246	0.91017
Lucilius - Book 6	149	0.93794
Lucilius - Book 7	127	0.92133
Lucilius - Book 8	78	0.93895
Lucilius - Book 9	287	0.80452
Lucilius - Book 10	43	0.88324
Lucilius - Book 11	113	0.95115
Lucilius - Book 12	38	0.80786
Lucilius - Book 13	58	0.98143
Lucilius - Book 14	115	0.98576
Lucilius - Book 15	156	0.90751
Lucilius - Book 16	79	0.89363
Lucilius - Book 17	62	0.94761
Lucilius - Book 18	8	0.74626
Lucilius - Book 19	58	0.98095
Lucilius - Book 20	92	0.85059
Lucilius - Book 22	19	0.56596
Lucilius - Book 23	3	0.50237
Lucilius - Book 25	0	0.00000
Lucilius - Book 26	498	0.55279
Lucilius - Book 27	278	0.92555

Lucilius - Book 28	246	0.97236
Lucilius - Book 29	494	0.82506
Lucilius - Book 30	555	0.75289

Appendix H - Math Sanity Check

In an effort to prove my mathematical methods are sound I have created this small appendix. Originally, I had programmed SVD from scratch in PHP. It worked well, but it was too slow for a front end GUI (graphical user-interface). Instead I used an open source library by Doug Rohde (SVDLIBC) based upon the SVDPACKC library that was written by Michael Berry, Theresa Do, Gavin O'Brien, Vijay Krishna and Sowmini Varadhan. This library can be downloaded from the following sites: http://tedlab.mit.edu/~dr/SVDLIBC/ or http://beta.septuagint.org/svdlibc.tgz. Doug Rohde originally programmed this library while at MIT. He currently works for Google. I wrote a method for PHP to talk to this library in order to perform all SVD operations.

We will outline a simple SVD example. Let us begin with a simple matrix.

2	1	4	1	4	5
0	1	1	2	1	2
3	1	1	1	3	4
5	4	4	3	3	2

We use Doug Rohde's library to decompose this matrix. Be aware that factoring using SVD results in Σ, U and V^T as described above. U and V^T can differ each time you factor your original matrix, however the eigenvalues in Σ stay the same. Even though U and V^T differ, they are geometrically similar to your original matrix.

Σ					
12.842	0.000	0.000	0.000	0.000	0.000
0.000	4.169	0.000	0.000	0.000	0.000
0.000	0.000	2.121	0.000	0.000	0.000
0.000	0.000	0.000	1.786	0.000	0.000

U			
0.5837570	0.2110590	0.4429620	0.6468900
-0.5665920	-0.1022480	-0.3096830	0.7567140
0.3241720	0.5908940	-0.7384730	0.0203493
-0.4828180	0.7719090	0.4031610	-0.0922181

V^T					
0.4462430	0.2978670	0.4342320	0.2639300	0.4528450	0.4988560
0.4128560	0.4912840	0.0835999	0.2852660	-0.2464470	-0.6626370
-0.6909340	0.1216480	0.5802120	0.3906730	-0.1257960	-0.0521230
-0.1216260	0.1810480	-0.6298870	0.6648400	-0.1268280	0.3123670

Next, we use our factored matrix to calculate a rank-4 approximation matrix.

Doc 1	Doc 2	Doc 3	Doc 4	Doc 5	Doc 6
1.99999	0.99999	3.99998	0.99999	3.99998	4.99999
-0.00000	1.00000	0.99999	1.99999	0.99999	1.99999
2.99999	1.00000	0.99999	0.99999	2.99999	3.99999
4.99999	3.99999	3.99998	2.99999	2.99999	1.99999

With our new matrix, we can calculate our Pearson correlations against each document vector (each column represents a synthetic and simplified document matrix). We will calculate Documents 2-6 against Document 1. In other words, we are trying to find which document is most similar to Document 1. We obtain the following coefficients.

Doc 1	Doc 2	Doc 3	Doc 4	Doc 5	Doc 6
1.00000	0.80064	0.55470	0.41812	0.57266	-0.05338

Below I have included the simple PHP function which calculates the Pearson coefficient between two document vectors.

```
####################################
function pearson_vector($v1,$v2)
####################################
	{
	$all_fields=count($v1);
	foreach ($v1 as $first)
		{
		$second=array_shift($v2);

		$sum_xy+=$first * $second;
		$sum_x+=$first;
		$sum_y+=$second;
		$sum_x_squared+=pow($first,2);
		$sum_y_squared+=pow($second,2);
		}
	return sprintf("%.5f",( ( $sum_xy - ( ( $sum_x * $sum_y ) / $all_fields) ) /
			sqrt( ( $sum_x_squared - ( pow($sum_x,2) / $all_fields ) ) *
			    ( $sum_y_squared - ( pow($sum_y,2) / $all_fields ) ) ) ) );
	}
```

Appendix I - Personal Pronoun Counts

Satire	Occurrences	Total Words
J14	34	1,524
J3	35	1,419
J6	41	3,085
P5	35	923
H1.6	47	676
H2.3	71	1,657