

Summer 2020

Calculating a Hero: Computational Analysis and Chivalry in Chaucer's the Canterbury Tales

Alexander Handley Humphreys

Follow this and additional works at: <https://scholarcommons.sc.edu/etd>



Part of the [English Language and Literature Commons](#)

Recommended Citation

Humphreys, A. H.(2020). *Calculating a Hero: Computational Analysis and Chivalry in Chaucer's the Canterbury Tales*. (Master's thesis). Retrieved from <https://scholarcommons.sc.edu/etd/6084>

This Open Access Thesis is brought to you by Scholar Commons. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact digres@mailbox.sc.edu.

CALCULATING A HERO: COMPUTATIONAL ANALYSIS AND CHIVALRY IN CHAUCER'S *THE
CANTERBURY TALES*

by

Alexander Handley Humphreys

Bachelor of Arts
University of South Carolina, 2011

Submitted in Partial Fulfillment of the Requirements

For the Degree of Master of Arts in

English

College of Arts and Sciences

University of South Carolina

2020

Accepted by:

Michael Gavin, Director of Thesis

Holly Crocker, Reader

Cheryl L. Addy, Vice Provost and Dean of the Graduate School

© Copyright by Alexander Handley Humphreys, 2020
All Rights Reserved.

DEDICATION

To my parents, who proved time and again that they would do anything to make this possible. To my brothers, who helped me believe in myself enough to attempt this work. To my friends, who stayed with me through all the times I was too busy working. And to Eris and Athena, who provided exactly the comfort and companionship I needed to get through this process. Thank you. I am humbled by your generosity of spirit and hope one day to be worthy of your kindness.

ACKNOWLEDGEMENTS

I would like to thank Dr. Holly Crocker for her assistance and patience with this project, and Dr. Michael Gavin for the same and for providing me with the inspiration for the project in the first place. Further, I would like to thank Oxford University for their work on the Oxford Text Archive, where I sourced the encoded text used for this project. Finally, my thanks to the University of Michigan for the digital edition of the Middle English Dictionary, which was also instrumental in this work.

ABSTRACT

This project aims to provide a basis by which distant reading techniques may be applied to Chaucer's *The Canterbury Tales*. The critical corpus is oddly devoid of studies examining these techniques as tools for understanding Chaucer's work. This paper endeavors to rectify this gap by demonstrating the kinds of insights made available by computational distant reading techniques as described by Johanna Drucker, Matthew Jockers and Jerome Bellegarda, among others. This study is founded on the belief that close reading and other forms of analysis needlessly exclude a broader view of the target work. It is not my intention in this study to entirely supplant close reading - I merely demonstrate the value that additional analytical tools can offer. Chapter 1, therefore, will function as an introduction, familiarizing the reader with these techniques and the study in a general sense. Chapter 2 will provide an overview of some of the previous literature on knighthood, both historically and in literature. Chapter 3 will focus on the implementation of these techniques through the R programming language and will lay out the methodology and specific objectives of the study. Chapter 4, finally, will present the results of the study and discuss how this type of work can be expanded to other works in the Middle English corpus to supplement, rather than replace, the traditional forms of analysis that have so long predominated the critical discourse.

TABLE OF CONTENTS

Dedication	iii
Acknowledgments.....	iv
Abstract.....	v
List of Tables	vii
List of Abbreviations	viii
Chapter 1: Computational Analysis and The Knights as Heroes.....	1
Chapter 2: Previous Scholarship.....	7
Chapter 3: Computational Techniques and <i>The Canterbury Tales</i>	20
Chapter 4: Analysis and Conclusions	32
Works Cited	49
Appendix A: R Source	59
Appendix B: Sample of Raw XML Data.....	65

LIST OF TABLES

Table 3.1 Word token frequencies in the <i>Tales</i>	27
Table 3.2 Characters-terms matrix from KnT.....	28
Table 3.3 Terms-terms matrix from the <i>Tales</i>	29
Table 3.4 Characters-terms matrix from WBT	30
Table 3.5 Characters-terms matrix from SqT	31
Table 4.1 Sentences matrix for “Arcite”	37
Table 4.2 Sentences matrix for “Emelye”	38
Table 4.3 Sentences matrix for “knyght”	39
Table 4.4 Sentences matrix for “Palamon”	40
Table 4.5 Sentences matrix for “Theseus”	41
Table 4.6 Sentences matrix for “Ypolita”	42
Table 4.7 Sentences matrix for “chivalrie”	43
Table 4.8 Sentences matrix for “curteisie”	44
Table 4.9 Sentences matrix for “fredom”	45
Table 4.10 Sentences matrix for “honour”	46
Table 4.11 Sentences matrix for “trouthe”	47
Table 4.12 Sentences matrix for “worthy”	48

LIST OF ABBREVIATIONS

CkP.....	The Cook's Prologue
CkT	The Cook's Tale
CIP.....	The Clerk's Prologue
CIT	The Clerk's Tale
CT	The Canterbury Tales
CYP.....	The Canon Yeoman's Prologue
CYT.....	The Canon's Yeoman's Tale
FranP	The Franklin's Prologue
FranT.....	The Franklin's Tale
FranW	The Franklin's Words to the Squire
FrP.....	The Friar's Prologue
FrT.....	The Friar's Tale
GP	The General Prologue
HI	The Host's Interruption of the Tale of Sir Thopas
HW	The Words of the Host to the Prioress
KnI	The Knight's Interruption of the Monk's Tale
KnT	The Knight's Tale
ManP	The Manciple's Prologue
ManT.....	The Manciple's Tale
Mel	The Tale of Melibee
MerE	The Merchant's Epilogue

MerP.....	The Merchant’s Prologue
MerT	The Merchant’s Tale
MilP.....	The Miller’s Prologue
MilT	The Miller’s Tale
MkP.....	The Monk’s Prologue
MkT.....	The Monk’s Tale
MLE	The Man of Law’s Epilogue
MLI.....	The Introduction to the Man of Law’s Tale
MLP	The Man of Law’s Prologue
MLT	The Man of Law’s Tale
NPE.....	The Nun’s Priest’s Epilogue
NPT	The Nun’s Priest’s Tale
PardI.....	The Introduction to the Pardoner’s Tale
PardP.....	The Pardoner’s Prologue
PardT.....	The Pardoner’s Tale
ParsP	The Parson’s Prologue
ParsT	The Parson’s Tale
PhyT.....	The Physician’s Tale
PrP.....	The Prioress’ Prologue
PrT.....	The Prioress’ Tale
RvP.....	The Reeve’s Prologue
RvT	The Reeve’s Tale
ShT	The Shipman’s Tale
SNP.....	The Second Nun’s Prologue
SNT.....	The Second Nun’s Tale

SqPThe Squire’s Prologue

SqTThe Squire’s Tale

SumP The Summoner’s Prologue

SumTThe Summoner’s Tale

ThPThe Prologue to The Tale of Sir Thopas

Th The Tale of Sir Thopas

WBP The Wife of Bath’s Prologue

WBT The Wife of Bath’s Tale

CHAPTER 1

COMPUTATIONAL ANALYSIS AND THE KNIGHTS AS HEROES

This study demonstrates a new method of examining *The Canterbury Tales* and how that method might be expanded to other works in the Middle English corpus. It will begin by introducing the concepts of distant reading and computational analysis as a supplement to traditional forms of literary analysis that have been applied in the past, most specifically close reading. The foundational premise for this study is that the close reading techniques commonly applied to individual works in the Middle English corpus can be supplemented by larger-scale computational analysis techniques described and developed by scholars such as Johanna Drucker, Matthew Jockers and Ted Underwood to provide a more complete understanding of the structure and composition of the work. That is, rather than being limited to an individual tale or other division, the techniques exemplified here can be applied to the text as a whole in order to support a traditional analysis of the divisions.

To that end, after applying these computational techniques to develop a model of the *Tales* text, we will examine the notion of knighthood as it is represented in the *Tales*. “The Knight’s Portrait” from the “General Prologue” will suffice to give us a baseline description of the Knight and his Tale presents us with additional depictions of the heroic in Theseus, Palamon and Arcite. We will see that, despite Chaucer’s overwhelmingly positive description of The Knight, he does not hold these traits as particularly important

in presenting his own version of classical heroes; rather, these characters are largely devoid of traditional heroic characteristics and seem to simply be presented because they comprise the most impressive of Chaucer's tales.

This introductory chapter will lay down a case for the application of computational analysis to *The Canterbury Tales*, as well as provide some background on the project itself. Chapter 2 consists of a review of previous scholarship relating to the concept of knighthood in general and as it is portrayed in *The Canterbury Tales* along with the fundamentals of the mathematical techniques we will employ later. Chapter 3 focuses on the research objectives, the specific methods used, as well as an explanation of the programming environment. Finally, Chapter 4 will explore the results of the computational analysis, examine the knightly characters and their relationships to the concepts of knighthood and conclude with some thoughts regarding the future application of these techniques to other works of Middle English literature.

In terms of the computational side of this study, it is perhaps best to offer a working definition of some of the terms employed, specifically a clarification of the term "division" as used here, along with the particular interpretation of "distant reading" that prevails. By "division", I mean any of the individual tales, along with the introductions, prologues and epilogues, etc. (herein also referred to as "surrounding texts") that appear in the main body text of *The Canterbury Tales*. With the exception of the portraits, which are included in the main text of the General Prologue, each major division has been included in the database as a separate data structure to facilitate a comparison of Chaucer's diction in each case. As regards the term "distant reading", while it is most often applied to a study of a multitude of works, in this case it serves to distinguish the

kind of analysis undertaken from a more traditional approach. For all intents and purposes in this study, “distant reading” and “computational analysis” are interchangeable.

In his original development of the concept of distant reading, Franco Moretti identified the necessarily narrow view afforded by traditional close reading methods. The hundreds, or often thousands, of works published in any given century or under any given literary paradigm are far too daunting a corpus for a detailed close reading analysis (Conjectures 55). Moretti, then, offers as a solution the idea of using computational analysis to bring these other works under review. While not a direct replacement for traditional analysis, computational analysis allows readers to consider all of the works from a given period, rather than simply – at the very best – only those works considered part of the literary canon.

In “Why Distant Reading Isn’t”, Johanna Drucker confirms some of Moretti’s claim and expands the scope of the discussion beyond purely literary texts. While Moretti seems to be more of a mind to have distant reading techniques supplant traditional forms of analysis as our primary method of engagement with a text, for Drucker this is an impossibility: “Processing is not reading. It is literal, automatic, and repetitive. Reading is ideational, hermeneutic, generative, and productive. Processing strives for accuracy, reading for leniency or transformation. No text-analysis program weeps...” (630). It is precisely these differences that allow the two perspectives of close and distant reading to supplement each other. Allowing a computer to do those things which it can in the process of a literary analysis frees the human reader to focus on the things they can do, without losing any of the interpretive power of either.

These techniques are often applied to large portions of any given corpus, famously in terms of thousands of books or their titles – as in Moretti’s “Style, Inc. Reflections on Seven Thousand Titles”. In addition, however, these techniques can also allow us to view an otherwise more manageable work – in this case, a single edition of *The Canterbury Tales* – in a new light. In the 2011 pamphlet “Network Theory, Plot Analysis”, for example, Moretti applies these techniques to individual works in a way that gives us new insight into their plot structure and the relative importance of different characters. Moretti’s analysis shows us, for example, a representation of the character connections found in *Hamlet*. Modeling these connections in terms of the proximity of individual characters’ lines, he finds that, as may be expected, Hamlet and Claudius act as hubs of connections between the lesser characters (5). This is not a new conclusion; rather, it simply confirms the conclusion of human readers with data that would be difficult, though not impossible, for a human to collect on their own. Similarly, this study will examine some basic statistical aspects of *The Canterbury Tales* that shed some light on its composition.

While of course it is possible to perform a detailed, close reading analysis of *The Canterbury Tales*, the more algorithmic (not to mention tedious) tasks that computational analysis allows make a quantitative analysis of the subject text all the more accessible. Further, even in the manual application of these tasks, the possibility of human error and the time investment necessary to accomplish them leaves much to be desired, as Jockers points out in *Macroanalysis*: “the sheer quantity of available data makes the traditional practice of close reading untenable as an exhaustive or definitive method of evidence gathering. Something important will inevitably be missed” (9). The kinds of analysis this

study employs take advantage of the specific strengths of a computational model to obviate the need for manual tabulation and calculation. Of course, the interpretation of those calculations still requires human intervention.

Critics of Moretti's work tend to focus on the deficiencies they perceive in the approach: that one is arguably not even reading the work, that a work turned into a statistical abstraction necessarily disregards its artistic integrity, and, more generally, that the move away from traditional literary analysis is in some abstract sense detrimental to the field (Trumpener 163). These critics, however, seem to view much of the relationship between close and distant reading techniques as one of replacement. Rather than trying to replace close reading techniques, computational analysis should be applied to fill in exactly the gaps that Moretti and Drucker describe. Again, I turn to Jockers for his assessment of the situation: "[t]he same argument [see above], however, may be leveled against the macroscale; from thirty thousand feet, something important will inevitably be missed. The two scales of analysis, therefore, should and need to coexist" (Macroanalysis 9). So, we must be careful, and the conclusions derived from this kind of study should be understood with the appropriate qualification with respect to a close reading analysis.

More than the computational concerns on the level of theory, this study is concerned with *applying* these techniques to a work of Middle English literature. We will begin with an analysis of the Knight as Chaucer describes him, as he is interpreted and how that analysis applies to the knights Palamon and Arcite in his Tale. That is, how do the words Chaucer uses to describe the Knight measure up to a modern understanding of a knight in Chaucer's time? Does that agree with the portrayal of the two knights in the

tale itself? Finally, do we need to reevaluate our understanding of heroism in this time as a consequence of the statistical characteristics that computational analysis reveals?

This study, then, undertakes some of the more labor-intensive tasks that could fall prey to human error, but which are readily performed by computers. In a broad sense, the project begins by parsing the entire text of *The Canterbury Tales* and separating it by category – that is, by tale or surrounding text division. It employs a variety of packages available for the R programming language to clean up the text, removing any numbers, punctuation or XML encoding tags that are present in the source data (Jockers, *Text Analysis* 121). This leaves us with a plain-text representation of the entirety of the *Tales*, separated by major headings which can then be further manipulated depending on the specific statistical methods desired. In the case of this study, by implementing a variety of computational techniques (Latent Semantic Mapping/Analysis [LSM or LSA] theory, Positive Pointwise Mutual Information [PPMI] and *k*-means clustering) to generate a computer-understandable sense of the meaning of Chaucer’s words, we will be able to fill in some of the gaps that more traditional forms of analysis leave in our examination of Chaucer’s Knight.

CHAPTER 2

PREVIOUS SCHOLARSHIP

As the purpose of this study is not to provide an especially far-reaching analysis of *The Canterbury Tales* as a whole, we will focus here on scholarship that deals with the three most pertinent topics: chivalry (and its associations with knighthood and noble virtue), for a foundation in the historical aspects of the concept; heroism in English literature in general; and, finally, the concepts of computational analysis used to re-examine the *Tales* from a new perspective.

Chivalry, Historical and Literary

While there is a great deal of scholarship on *The Canterbury Tales* over the intervening centuries, most important for this study are those critics who focused at one time or another on the Knight and his son, their place in medieval society, and the peculiar and often misunderstood concepts of “chivalry” or “knighthood” both in history and as presented by Chaucer.

Beginning with Charles Mills’ 1826 *The History of Chivalry*, we see numerous accounts of real-life chivalric conflicts and their resolutions for a bit of background. Indeed, it is obvious from these examples that historical knights were, frequently, a necessarily violent bunch. Fully invested in their codes of behavior, knights such as John de Visconti and Thomas de la Marche are described putting their lives on the line over a

simple matter of honor. The system is so ingrained in these men that, while willing to fight to the death over this matter, they were also willing to wait during the travel time between Armenia and England to resolve the question in front of the king:

The judgment was, that they should carry letter importing their cause fully and clearly from the said Christian princes unto King Edward of England, and should submit themselves to be tried by combat before him, as the most worthy and honorable prince in all Christendom; they swearing to remain as perfect friends until that time (23).

Even outside of the role of a knight as it applies to full-scale combat, the chivalric ideal reinforces the primacy of valor. Of course, that ideal cannot constrain real life too much; the trial by combat leaves the elements of a tournament behind and devolves into a hand-to-hand brawl (25).

Further, the exploits of Sir Walter of Manny show the value of particularly conspicuous valor. The foremost of his adventures, according to Mills, was an attempt to rescue “two brother-knights, whom an uncourteous cavalier, called Sir Loyes of Spain, had condemned to death” (29). Accomplishing a task that two knights who are themselves distinguished could not achieve for themselves would certainly demonstrate his prowess. He’s described storming an enemy encampment (33), successfully defending a castle against a sieging army (36), of impressing the chivalrous King of France with his demeanor (38), and causing the King of England to spare the majority of the inhabitants of Calais during a siege (39). Manny’s accomplishments in this regard are presented as

being of particular note, and in comparison to other knights either exhibiting only slightly lesser virtue or of substantially higher noble status.

Chapter II, “Progress of Chivalry in Great Britain, from the reign of Richard II to that of Henry VIII”, focuses in part on the historical context of the concept of chivalry in the time of Chaucer. While Chaucer’s Knight is certainly informed by Chaucer’s conception of knighthood in his experience, we are well served by familiarity with what Chaucer’s audience would have understood a knight to be. Mills describes a chivalric tradition in decline, largely from the failures of Richard II and his immediate predecessors: “[Richard II] had neither spirit nor ambition to recover the possessions which had been wrested from the crown during the illness of his father, the Black Prince, and the imbecility of his grandfather, Edward III” (64-65). Richard’s lack of valor seems to have directly contributed, according to his contemporaries at least, to the failure of the chivalric ideal (66). With that ideal as expressed in literature on the one hand and the real-life examples that Chaucer would have been exposed to in the other, there’s little short of actual villainy that we shouldn’t expect somewhere in the *Tales*.

D. W. Robertson’s *A Preface to Chaucer* provides a guide to many aspects of Medieval literature in general and, of course, Chaucer’s style in particular. In the chapter “Late Medieval Style”, Robertson asserts that, contrary to what we might expect, the Knight, Palamon and Arcite are not Chaucer’s primary heroic figures:

The “heroes” of *The Canterbury Tales* are Duke Theseus, Constance, Griselda and St. Cecelia – all figures who are distinguished by virtue rather than by heroic

action. It is true that Theseus is a mighty conqueror, but his physical heroism is passed over lightly and his wisdom and mercy are stressed instead (284-285).

It is not his classically heroic traits, those he might share with Achilles or Aeneas, that center him in this frame. Instead, for Robertson, it is Theseus' more Christian virtues that are important. Citing *Troilus* and *Alceste* as examples, Robertson asserts that in most, if not all of Chaucer's works, he "neither uses outward heroism as a symbol for spiritual heroism nor confuses the two". Like Mills, Robertson believes that much of Chaucer's skepticism for more traditional heroic representation stems from the chivalric failures of English kings: "an old Edward doting on an unworthy woman and a young Richard who could not always achieve those principles of chivalry which he so ardently admired may have made this distinction acutely apparent in Chaucer's lifetime" (285). This division between external displays of heroism and the internal, spiritual virtue is fundamental to understanding the function of a knight, as the bearer of "two swords proper to defense against the various enemies of man ... [t]he knight should gird on the external one to keep temporal peace safe from violence, and the internal one, which is the sword of the Word of God, to restore peace to his own breast" (174). That is, while engaged in a violent way of life, a knight required that his Christian faith protect him from becoming a source of the very violence he was sworn to prevent.

In *Chapters on Chaucer*, Kemp Malone describes the additional attention more recent (20th Century) scholars have paid to the pilgrims as objects of study, claiming that it wasn't until recently that "the men of learning, if not the general reading public, began to pick them [the pilgrims] to pieces" (163). Before that, it seems from Malone's perspective, the pilgrims were simply taken at face value, with no need for extraordinary

focus on their characterization or interactions. He describes the generally unique nature of each pilgrim, noting exceptions in the vagueness of the burgesses, the fact that there are technically two carpenters, and the mysterious pack of priests following the Prioress more as outliers. Malone comments here on the exceptional nature of each of the primary pilgrims, along with the ironic commentary that “[t]he superlative quality of Chaucer’s pilgrims is beautifully brought out in the passage which I have just quoted. The explanation given is less satisfying”. That is, that the pilgrims are, by and large, perfect or exceptional in their characterization is hardly exceptional for literature of Chaucer’s era: “From time immemorial it has been the custom in story-telling to make the characters heroic, larger than life, extraordinary ... [n]obody in the fourteenth century would have thought of doing otherwise” (167).

For our purposes, naturally, we shall focus on the “perfection” of the Knight. Here, again, we must look past a technical definition of “perfection” and instead look at his characterization. As Malone puts it, “Chaucer calls him a perfect knight, not a perfect man, and the distinction is important. His perfections are those of knighthood. For that very reason his weakness in the literary realm does not matter much” (173). The Knight is not a Christ-figure, perfect in all ways and immune to the weaknesses of humanity. Instead, he is the near-perfect example of his type, with an heroic history and (with one exception) possessed of a gentle disposition and impeccable courtly manners. That he fails even in this narrower view of perfection is yet further evidence of his humanity.

In later chapters, Malone continues his examination of the Knight’s characterization. Malone is convinced that as a result of the possible pre-existing nature of “The Knight’s Tale”, Chaucer “paid little heed to his characterization of the knight

when he gave him a tale to tell ... the knight was given this particular tale in virtue of his knighthood, his characteristics as an individual playing only a minor part in the selection of a tale for him” (231). Most of the other pilgrims either seem to be using their tales to represent themselves – Alisoun, in particular, uses her tale to act as a sort of defense for her behavior – or in an effort to engage in a bit of banter with another pilgrim. The difference with the Knight is somewhat striking, then. That the reader does not become distracted by how disconnected “The Knight’s Tale” is from its teller and the general pattern of the rest of the stories deserves some attention.

In his discussion of the Prioress, Malone seems to think there is a mechanism at work in disguising how unfitting “The Prioress’ Tale” is to the description given her in the General Prologue: “The story ... is given to the prioress because a nun seems a suitable person to tell it. The earlier characterization ... is dropped. It is several thousand lines away, and the reader or hearer has probably forgotten it by this time” (219). “The Knight’s Tale” being placed so near the beginning of the *Tales* may serve to camouflage a similar ill fit between teller and tale. By the time the reader or listener realizes there may be a strong correlation between many of the pilgrims and tales, they likely have forgotten that the exceptionally heroic (and violent) Knight tells a love story, albeit one that ends in what amounts to a battle.

Johan Huizinga’s *The Waning of the Middle Ages* gives a solid in-depth accounting of the meaning that chivalry as an ideal held in the medieval world. Indeed, : “Medieval thought in general was saturated in every part with the conceptions of the Christian faith ... all those who lived in the circles of court or castle ... their whole system of ideas was permeated by the fiction that chivalry ruled the world” (67). Indeed,

we see that knighthood and chivalry were deeply intertwined with Christian mythology and symbolism when Huizinga declares the archangel Michael as the progenitor of feats of arms as a model for human behavior. He is not, of course equating the two. Huizinga draws clear distinctions between a properly Christian, pious life and the violence of knighthood.

Huizinga views chivalry as a belief system that, much like many religions, offered a human-understandable explanation for the complex workings of the world. In this case, Huizinga argues that “chivalry constituted for these authors a sort of magic key, by the aid of which they explained to themselves the motives of politics and of history” (68). Rather than explaining natural phenomena, the religion of chivalry allowed authors and audiences alike the ability to pretend that what they saw around them had a purpose. There is every opportunity for chivalry to provide a veneer of reasonability to a world full of greed and violence if one can simply explain it as violence done in the name of the right cause.

Huizinga describes a complicated interplay between the real world and the popular conceptions of ideal chivalry. We’ve seen attention paid earlier to the dichotomy of Christian piety as it relates to chivalry (famously pacifist in theory, famously warlike in practice), and Huizinga highlights this even further. Unlike the ideals of Christian piety, sourced from God himself and infusing the natural order of things, ideal chivalric virtue is more mundane:

medieval thought did not permit ideal form of noble life, independent of religion.

For this reason piety and virtue have to be the essence of a knight’s life. Chivalry,

however, will always fall short of this ethical function ... [f]or the source of the chivalrous idea is pride aspiring to beauty (69).

Excellence as it would have been viewed in Plato's time was simply antithetical to the life of a pious knight, and yet inherent in the quest for chivalric virtue. How, then, does one reconcile these very disparate models of behavior? For Huizinga, the answer is simply that "[t]he life of a knight is an imitation" (71). Princes and knights alike, in Huizinga's view, were merely attempting to present themselves as achieving the chivalric (or Christian, for that matter) ideal. The vows these men swear take a form "akin to purely religious vows, serving to accentuate or to fix a lofty moral aspiration" (89). That is, it is unlikely to be the case that anyone could actually expect a strong and perfect adherence to the letter of the vows. Such an expectation could hardly fail to disappoint.

In *Chivalry and Violence in Medieval Europe* Richard Kaeuper describes a high-level division of medieval society as "threefold: the imagined world divided into those who fight, those who pray, and those who work" (8). Being careful to note that, contrary to what one might expect their claims to be, the fighting did not always have a noble cause, Kaeuper instead focuses on the more brutal characteristics that these supposedly chivalrous knights share with all warriors: "[t]hey fought each other as enthusiastically as any common foe; perhaps even more often they brought violence to villagers, clerics, townspeople, and merchants" (8). In chapter 1, he focuses on a series of critics (Orderic Vitalis, Suger and Galbert of Bruges) with largely positive opinions regarding the need for order and a subsequent willingness to overlook the violence that that order implies, pointing out in the first case that "Orderic is no pacifist. Violence in the right cause, carried out by the proper people can cause him to wax eloquent" (13), in the second that

“Abbot Suger was not second to Orderic in his admiration for royal agents of imposed peace, nor in his belief that they might act with force in the interests of order” (16), and finally, regarding Galbert of Bruges, “[h]owever much his point of view might differ in detail from our monastic writers ... his account dovetails with their emphasis on ... the need for a strong authority figure to repress violence and secure peace” (18). All three of Kaeuper’s initial historical touchstones agree that a lack of strong authority leads to disorder, and that violence may be required to restore peace. This is the central fiction behind the attempt to legitimize the violence of knights.

We see this acceptance of the need for and value of violence through Kaeuper’s analysis of chivalric literature. Of this particular genre, he notes “[b]elief in the right kind of violence carried out vigorously by the right people is a cornerstone of this literature” (22) - there is no room in chivalric or heroic literature for a pacifist knight. Relaying a scene from *Perlesvaus*, predating *The Canterbury Tales* by some 150-odd years, the difference between the “right” and “wrong” kind of violence or people is in stark relief. The ordered violence of Arthur and his knights is contrasted with the chaotic violence of the “demonic knights”: “[t]hough the maiden warns Lancelot not to step outside the protective magic of his circle, with characteristic valour he attacks the knights ... the evil is defeated” (23). There is no substantive difference between the violence committed by either side; rather, it is in the re-establishment of order that the difference becomes clear.

Kaeuper goes on to describe the connection between chivalric virtue and prowess in battle, a characteristic we’ve seen in Chaucer’s description of the Knight already. In fact, he designates physical prowess as chief among the things that defines a knight:

“[n]ot simply one quality among others in a list of virtues, prowess often stands as a one-

word definition of chivalry in these texts” (135). Given the primacy Kaeuper gives to prowess in determining the worthiness of a knight, we can take the Knight’s illustrious battle record as the only proof necessary of his virtue; his courtliness and culture can, in light of this assessment, be viewed as merely an intensifier. It is not simply for our understanding of knights in this tradition, however, that we should put so much emphasis on prowess: “both imaginative literature and the historical accounts of their lives picture knights enjoying a privileged practice of violence; it suggests that they found in their exhilarating and fulfilling fighting the key to identity”. That is, prowess in battle, victory, and the confinement of those things to an established form of *knightly violence* were key to the knights’ self-conception. As he described the contrast between Arthur’s cohort and the demonic knights in *Perlesvaus*, Kaeuper asserts that the “ideal chivalric figure is not, of course, a latter-day Viking berserker” (143). There is a fine line to be drawn here. It is not enough to be larger or stronger than your opponent. A proper knight of good virtue fights his opponent on a largely equal footing, but we must not overlook the bloodlust that fuels their prowess: “[t]o read much chivalric literature is to find admired knights regularly feeling rage as they fight; their blood boils; when honour is challenged, they nearly lose their minds”. It is not unusual for a truly chivalrous knight to be overcome by this bloodlust – it is simply the mark of an exceptional knight to be able to restore his own wits through clarity of mind (144-145). Kaeuper is also careful to note the disparity of real-life combat (requiring planning and leadership by a skilled general) and the kinds of single combats often portrayed in chivalric literature (146). Much like the failure of Mills’ knights to maintain a purely orderly duel, the realities of war often did not allow for the fancies of romance authors.

For Nigel Saul, military adventures in service of the king were tied largely to monetary concerns. In *Chivalry in Medieval England*, Saul describes the arrangements made between the king and his various captains to provide soldiers for campaign. While we cannot assume that no knights were motivated on the basis of principle, serving their king for glory or the advancement of Christendom, Saul is clear on this point: “The commercial character of these indentures ... is indicated by the careful attention given in them to all matters financial ... [t]he sheer popularity of the indenture system affords clear evidence that those involved in it saw military services as a source of profit” (120). Saul continues on to describe the sources of profit that a knight could expect to exploit: direct acquisition of loot on campaign, ransoms paid to return prisoners, and the income one could expect from being granted lands in the newly-conquered territories. Materiel such as arms, armor and horses could be quite profitable indeed, but we see that the capture of prisoners (such as one Guy taken prisoner by Sir Walter Manny) could provide a windfall in excess of many years’ worth of a knight’s usual income (122). If not glory, duty or God, certainly the fortune that could be gained on campaign was a powerful motivator. A single high-value prisoner could see a knight living in relative comfort for the rest of his days, or provide him such arms, armor and luxuries as to simulate glory among his peers.

Of course, Saul does not argue that this is the sole motivator, or that it was even especially commonplace for an English knight to cover himself in wealth. This is to say nothing of the common soldiery serving under him. Saul is careful to note that the opportunity for prisoner-taking was rarer than might be expected: “the number of battles at which valuable prisoners were taken was relatively small ... [a]t Crécy the French king

... had ordered that no quarter be given, and Edward III in retaliation had ordered the same” (125). It is, instead, the skirmishes and minor battles that seemed to produce the most ransom money, because the risk of knights and soldiers being distracted was relatively low. While there were outliers on both ends of the profit spectrum, it seems that the majority of knights probably found themselves breaking even, barely balancing the costs of their outfitting and adventuring with the occasional bit of looting or minor ransom (126). So, while the extravagant sums acquired by some (such as Walter Manny) no doubt provided motivation for a knight, we cannot think that it was the only one, or that any given knight had only one motivation himself:

If this view of military service as rooted essentially in a quest for honour appears at odds with the apparent materialism of some of the knightly class ... it may have been the case that honour was a quality esteemed even by those who were moved principally by a quest for booty and profit. In the medieval view of the world profit and honour might not be opposites; they could go together – the accumulation of profit might be a mark of honour, in some circumstances even a source of honour (127-128).

Proving one’s valor and prowess, and by extension one’s honor and virtue, by displaying the wealth taken from one’s enemies in service of the king was a potent way to ensure future opportunity. Fame and fortune go hand-in-hand, and a knight who managed to capture great quantities of wealth would have been famous indeed.

Again, we must be careful to assume too much from a small number of examples. We can also not assume that the knightly class was universally renowned by the English people. Many saw the knights, as we’ve seen before, as sources of violence that were

counter to the will of God. Some others, that the knights were essentially not violent enough - that “they were becoming too soft” (130). It is these failings, either on the extreme end of violence or its opposite, that many English commentators blamed for English failures in war (131). Along with the growing civilian nature of a knight, especially those knights employed directly by the king himself, came the standardization of codes of behavior on campaign. Rather than focusing so much on knights as individual beacons of heroism – or sources of violence, questers for treasure, etc. – these codes were intended to ensure that knights served the good of the cause rather than their own self-interests. This change served to revitalize the reputation of the knightly class, binding them by a set of rules that would restrain them from wanton pillaging and random violence, with the usual sorts of punishment for violation of those rules (134).

In later chapters, Saul focuses on the differences between the chivalric ideal and the reality of knighthood: “[o]n the one hand it sustained the ideal of the perfect knight – the knight who fought to maintain order, defend the faith and protect the weak. On the other ... it fed the oppressions of a predatory knighthood” (193). Part of the code binding the behavior of knights included a call for violence in defence of one’s honor, or the honor of one’s family. This, naturally, lead to instances of largely unjustified violence – violence perpetrated not on the order of the king, or for the glory of God, but at the whim of the knight who had been slighted. It is not hard to claim personal honor as a justification for violence, as only the knight himself would know what satisfaction his honor required. We may be tempted to assign blame for this behavior to the types of literature these knights consumed, but Saul is careful to note that there is little evidence to

suggest that knights modeled their behavior wholesale on the “aggression of the chivalric ethic” (196).

In *Chaucer's Troilus and Criseyde and The Knight's Tale: Fictions Used*, Frieda Penninger highlights, again, the chivalric characteristics Chaucer assigns to the Knight, noting in particular that “he is the first pilgrim presented and the one elected to tell the first tale” as proof of his status among the Canterbury pilgrims. While his status is certain, she is not so convinced of his virtue, as she points to the “ambiguous” and “perhaps directly satirical” interpretations of other scholars, and emphasizes that “The Knight is not entirely occupied with serving God and country” (81). He is a man of his trade, such as it is, and like everyone else must provide for his sustenance and the maintenance of that trade. For many people gifted with his skillset, the life of a mercenary is a suitable source of income.

Penninger pays particular attention to the Knight’s apparent attitude about the frame story and its tale-telling contest. He cannot be anything but wholly certain in his impending victory. His confidence in his ability to win the contest seems natural from a warrior with as much experience as he: “Ful worthy was he in his lords were, / And therto hadde he riden, no man ferre,” (Chaucer GP 47-48). A knight with such experience and so much success behind him would hardly balk at the challenge afforded by opponents of decidedly lower station, even if it is in a contest reliant on his one failing as a perfect, noble knight. Indeed, a trip to Canterbury would seem like a stroll down the street for someone who had been at “Alisaundre”, “grenade”, and “algezir”, among other battlefields (51-57). So “parfit” and “gentil” a knight has quite a daunting real-world

standard of conspicuous valor to live up to (72), and of course any literate audience would expect him to exceed that standard.

Penninger goes on to describe in particular the Knight's complicated characterization. For her, there is a great deal of conflict between the Knight's clear accomplishments as a warrior and his tale: "the Knight chooses to tell a love story rather than a story of military conquest, a surprising choice which is underscored by the fact that when the Knight could put his professional expertise to use in developing the military side of his story, he generally does not" (84). Similar conflict arises in her discussion of the Tale, where we see both the knightly oaths used "as a reason for fidelity", while the knights subsequently break those oaths for their own self-interest (85). Indeed, as the Knight tells of Arcite's death, there is a conflict between the gory and overly-specific description of his wounds and the Knight's characterization in his portrait:

The brutality of the Knight's description of Arcite's wounds and the casualness of his dismissal of the dead man ... have occasioned some comment as to what they may imply about the Knight's callousness towards injury and suffering and his hesitation to enter into a discussion of religion, both points which may suggest that he may not be entirely perfect and gentle but which also remind the reader of the ugliness of violent death and of the problem of soul's welfare for those who live and die by the sword. Not only the tales which *quite* the Knight but *The Knight's Tale* itself diminish knighthood (87).

If it was Chaucer's goal to provide us with an idealized vision of knightly virtue, then it is not in the Knight or his tale that we can find it.

As we can see, the literary concept of knighthood and its reflection in the real world bear little similarity to each other. Moreover, the complicated nature of their interaction leads to some often-conflicting behaviors in individuals who are expected to inhabit the ideal. Knights are supposed to stand for God, honor, justice, truth and the like; however, in order to do so they must often also commit horrific acts of violence in the name of their lord. In order to reconcile these conflicts, we will have to look at the computational techniques this study employs.

Computational Analysis

Matthew Jockers's *Text Analysis With R* is, as the name might suggest, a primer for interacting with the R programming (and RStudio) interface. The specifics of the R programming language are beyond the scope of this analysis; it merely serves as a method by which to implement the theories set out by others. Working through the exercises Jockers presents will provide a strong foundation for manipulating the data and getting some rudimentary statistical measurements out of a text. For the more rigorous measurements, however, we will have to look elsewhere. Still, while we cannot proceed with just the techniques Jockers provides, it is invaluable for this study and proved to be superior to other guides with similar goals.

We can find the desired rigor in Landauer, et al.'s *Handbook of Latent Semantic Analysis*. In its first chapter, "LSA as a Theory of Meaning", Thomas Landauer argues for the adoption of LSA in machine learning as a method of representing a "humanlike" understanding of language in a computer:

LSA is a computational model that does many humanlike things with language. The following are but a few: After autonomous learning from a large body of representative text, it scores well into the high school student range on a standardized multiple-choice vocabulary test; used alone to rate the adequacy of content of expository essays ... estimated in more than one way, it shares 85%-95% as much information with expert human readers as two human readers share with each other... (Landauer 5)

These are, of course, very impressive accomplishments for a computer model of language. For this reason, LSA (and as we'll see, its extension LSM) has become a leading theory of computer learning, and as such is ideally suited to the kinds of analysis this study undertakes.

Landauer further elaborates on this initial argument, offering examples of the kinds of measurements that LSA allows one to take. He describes here the concept of cosine similarity that is foundational to the mathematics underlying this study (17-18). Specifically, cosine similarity is the main characteristic of word-word and word-sentence pairs that we will explore in later chapters. We will facilitate a meaningful (and computationally feasible) implementation of cosine similarity using PPMI and other algorithms designed for this purpose.

In "Latent Semantic Mapping", Jerome Bellegarda lays out the mathematical foundations for this theory as a method of expanding on the existing LSA theory by applying similarity calculations to "compositions" (otherwise called "documents" elsewhere in this study. Whereas the standard LSA theory only accounts for the

collocation of terms to other terms, Bellegarda's LSM attempts to include the semantic information of a term's context in the calculation. This study adapts Bellegarda's even further, generally considering as the context only the sentence in which these terms appear instead of a whole composition.

Peter Turney and Patrick Pantel's "From Frequency to Meaning: Vector Space Models of Semantics" provides many of the other major components required to perform this analysis. They describe, step by step, the process of building a Vector Space Model (Turney and Pantel 153-155), how to manipulate the resulting matrices to conform to the requirements of the PPMI algorithm and provide a formal definition of the mathematics underlying the process (156-157). They describe a variety of weighting methods such as frequency analysis, tf-idf and the standard Pointwise Mutual Information (PMI) function, noting that each subsequent method is an improvement over the previous one (156-157). We can safely ignore these earlier methods in favor of PMI and PPMI. In essence, PMI calculates the probability that a given term occurs in a given context, given as a range from -1 (completely unrelated) to 1 (perfectly correlated). PPMI modifies the standard PMI method by replacing all values less than 0 with 0 to simplify the matrix calculations required later. It happens that this modified PMI method "performs better than a wide variety of other weighting approaches when measuring semantic similarity with word-context matrices" (157), which is exactly the kind of measurement this study focuses on.

Further simplification of the matrix is described later, where Turney and Pantel discuss the use of Singular Value Decomposition (SVD) in terms of its varied effect on the computation: as a representation of latent meaning, as a method of noise reduction in the matrix, as a measure of high-order co-occurrence of terms, and as a method of

sparsity reduction (that is, making the matrix more “dense” by removing extraneous values) (158-160).

However, we should not make the mistake of assuming that a data-driven analysis such as this is without its biases. Human failings will always creep into literary analysis, as Johanna Drucker points out in “Why Distant Reading Isn’t”:

[d]esigning a text-analysis program is necessarily an interpretative act, not a mechanical one, even if running the program becomes mechanistic. The contrast between machine processing and human cognition remains, but the automated methods are also fraught with cultural, historical, and other prejudices built into their design. The methods are based on epistemological assumptions that get translated into metrics and always operate only within the limits of current technical capabilities (631).

The goal cannot be to remove these biases from the analysis. Such a task would be patently impossible to achieve. We must, instead, mitigate the effects of our biases on the interpretation by qualifying our conclusions, removing bias where we can, and noting it where we cannot.

CHAPTER 3

COMPUTATIONAL TECHNIQUES AND THE CANTERBURY TALES

While previous chapters dealt with the general structure of this study and the prior scholarship surrounding Chaucer’s seemingly heroic Knight, this chapter will focus on the specifics of the process. Matthew Jockers’ *Text Analysis with R* provides a solid foundation for the use of the R programming language in doing this kind of work. Using this as a base of reference, we can construct a quantitative analysis of the lexical profile of *The Canterbury Tales*.

The statistical characteristics that this study will investigate, cosine similarity and PPMI primary among them, form the basis for a more complete understanding of the *Tales*. For reference in this chapter, the R scripts used to perform the various computations are included in Appendix A, while a sample of the raw XML data is included in Appendix B. The code has been commented to give some insight into what each block of instructions is intended to do.

The basic statistical characteristics form the foundation, but the primary analytical paradigm used in this study is LSA, a process which seeks to discover the underlying semantic relationship between a set of *terms* (usually individual words or search queries) and a corpus of *documents* (in the case of this study, sentences within the *Tales*). We can examine, for example, the terms most semantically related to an individual character (“Arcite”), the sentences most associated with an individual term (“Curteisie”), and even

construct a topic model of the most important terms in each individual tale (see Chapter 4).

The first step in this project is to locate a suitable, encoded edition of *The Canterbury Tales*. An appropriate edition of the text must satisfy some basic requirements. In order to facilitate the discovery of word tokens and their separation into various classes, each type of surrounding text (prologues, epilogues, introductions, portraits) as well as each individual tale must be clearly notated in the text source. Further, the text of each tale must be readily separated from the XML encoding tags to facilitate the calculations.

The University of Oxford's Oxford Text Archive (OTA) provides a variety of digitized editions in a range of formats. For the purposes of this study, the digitized edition of Fred Norris Robinson's *The Works of Geoffrey Chaucer* satisfies the requirements necessary to break the text up into its component parts with ease. In the general case, one also often must deal with peculiarities of orthography that are often present in Middle English texts. In this particular case, Robinson's edition has removed some of the need for replacing 'f' with 's' or a significant proportion of 'u' with 'v', for example. An ideal source text would have no page numbers or other extraneous content outside of an XML tag, but in the case of the source used for this study, I have had to make some modifications or allowances.

Tokenizing the text is the next major step in conditioning the data. Examining the XML source text for Robinson's edition of the *Tales*, we find that not only is the text divided into the traditional Fragments and Groups, which we will ignore for the purposes of this study, but also by division type and individual lines:

```

<div n="1" type="fragment">
<head>Group 1</head>
<div n="GP" type="prologue">
<head>The General Prologue</head>
<l n="1">Whan that aprill with his shoures soote </l>
<l n="2">The droghte of march hath perced to the roote, </l>
<l n="3">And bathed every veyne in swich licour </l>
...
<div xml:id="squire" type="portrait">
<head>The Squire's Portrait</head>
<l n="79">With hym ther was his sone, a yong squier, </l>
<l n="80">A lovyere and a lusty bachelor, </l>

```

We must proceed by identifying those labels in the XML that we will *not* be ignoring for the study. The character portraits, introductions, prologues, tales and epilogues all contain valuable information that can contribute to this tokenization, whereas pulling all of the individual lines out of the text will result in a mass of text that can no longer be separated into these larger divisions. So, for example, we place everything identified as a division of interest into its own character vector and identify their headers as the first “h3” element in each index of that vector. One of the peculiarities of this operation in R is that many of the elements may be NULL or NA rather than containing any part of the actual Tales text. We then find all entries in the vector that are not NA and use that to filter out the meaningless vector indices. This leaves us with a character vector whose members are

whole divisions, but which still contains a significant amount of XML encoding and other text that we need to filter out before we arrive at the pure word tokens that we'll be dealing with:

```
tales = xml_find_all(read_html("Oxford_Canterbury_Tales.html"),
    "-//div[@class='tale' or @class='prologue' or @class='introduction' or
    @class='epilogue']"# or @class='portrait']")
tales_vocab = xml_text(tales)
tales_txt = xml_text(tales)
```

To continue the tokenization process, we apply a series of operations on these vectors which removes excess whitespace and punctuation, splits the text into a vector whose elements are, not whole tales, but individual words as they appear in the *Tales*, reverts capital letters to lower-case (as the system will make a distinction between “the” and “The”, for example), and removes all vector elements that are now empty as a result of these operations (see Appendix B for the full text of the script).

Examination of the character vector at this stage reveals that the elements are either real words, numbers (an artifact of the original text encoding), or the letter ‘s’, due to the script stripping out the apostrophes in, for example, “Knight’s”. It is a simple matter to remove the few remaining extraneous vector elements. This will prove to be a moot point once we filter out each so-called “stop word”, trivial words like “a”, “and”, “the”, etc. that do not contribute significantly to the meaning of the text. After that operation, the first 15 elements of this vector are now "general", "prologue", "april", "shoures", "soote", "droghte", "march", "perced", "roote", "bathed", "veyne", "licour", "vertu", "engendred" and "flour".

R provides a base function, *table()*, which takes a character vector and returns a list of all unique tokens and the number of times each appears in the vector. We could use this function to compare the lengths of the word token vectors both with and without stop words removed to see the proportion of non-trivial words contained in the text. These are, of course, only the most basic of calculations that we can employ. Still, *table()* is useful for

Jerome Bellegarda describes a process, Latent Semantic Mapping, by which linear algebraic techniques can be applied to these data structures to develop mathematically rigorous descriptions of the structure of the text (Bellegarda 71-73). To begin, we construct a matrix in which each row is a word token, each column is a division of the text, and each cell therein is the count of the frequency of that word in that division (see Table 3.1 for a sample of this matrix). In addition to this mere tabulation of appearances of a word, we need to construct a matrix where, rather than the columns being divisions of the text, they represent the context of each appearance. We therefore divide the whole *Tales* text into its individual sentences and assign those sentences to the columns of the matrix. As with the previous matrix, each cell represents the frequency of appearances of a word token in each context.

These elements are still not necessarily significant; it will be our next task to identify those terms which are. This study uses the PPMI technique to weight the terms based on the likelihood that a random instance of a term is correlated to any given sentence. Even in the case of an individual work such as *The Canterbury Tales*, the matrices quickly become computationally cumbersome to work with; the number of unique word tokens (11,986 with stop words and other garbage removed) and the number

of sentences (8,551) means that we have constructed a matrix consisting of over 102 million elements. We can use SVD as described by Turney and Pantel to reduce this matrix to a more manageable size by attempting to filter out the noise inherent in a statistical sampling of such a random set of elements as word tokens in natural language. This study uses a modified implementation of SVD, the Implicitly Restarted Lanczos Bidiagonalization Algorithm (IRLBA) to aid in the computability of the decomposed matrix.

Next among the major computational tasks is to compute the cosine similarity of terms and sentences of interest to identify similar terms and sentences in the text. Given two vectors, in this case, the PPMI-weighted value vectors for two words in the text, cosine similarity analysis gives us a measure of their “distance” in the vector space expressed as a number between 0 (completely dissimilar) and 1 (identical). These scores are not, of course, necessarily significant unless we have some idea of what they mean on the continuum of 0 to 1. In general, we can regard any score below 0.3 to be essentially uncorrelated, anything between 0.3 and 0.6 as weakly correlated, and anything above 0.6 as a strong correlation.

We can apply this calculation at will to any combination of terms and/or sentences to find those elements of the matrix which are most semantically similar. For example, calculating the ten most-similar terms to the word “theseus” which aren’t “theseus” itself gives us the following list and associated values: "duc" (0.896747), "howleth" (0.784545), "gliterynge" (0.753909), "atthenes" (0.751701), "fightyng" (0.712322), "perotheus" (0.693676), "arcite" (0.682116), "palamon" (0.677625), "rit" (0.675698), "martireth" (0.671662) and "ypolita" (0.664422). As should be expected,

among the things Theseus is associated most strongly with are his city, the concept of war, his friend, the primary protagonists of the tale and his wife.

Finally, we can construct topic models for each tale (or the *Tales* as a whole) using the k-means clustering algorithm initially developed by J. A. Hartigan and M. A. Wong and as implemented as the default option in R. This topic modeling scheme allows us to programmatically determine the terms most semantically significant to a list of topics chosen from the word tokens of the text. Due to the random nature of the seeding process, subsequent iterations of the k-means algorithm will show different results; however, each iteration of the algorithm supplies us with a set of k topics and an arbitrary number of related terms and their semantic similarity scores. Table 3.2 shows the results of this similarity scoring operation on a list of terms from “The Knight’s Tale”. Tables 3.3-3.5 show, respectively the scores for terms across the *Tales* as a whole, “The Wife of Bath’s Tale” and “The Squire’s Tale” for further illustration. We will explore the implications of these scores further in Chapter 4.

Table 3.1 Word token frequencies in the *Tales*

	GP	KnT	MilP	MilT	RvP	RvT	...	CYP	CYT	ManP	ManT	ParsP	ParsT
general	1	3	0	0	0	0	...	0	0	0	0	0	4
prologue	1	0	1	0	1	0	...	1	0	1	0	1	0
april	1	0	0	0	0	0	...	0	0	0	0	0	0
shoures	1	0	0	1	0	0	...	0	0	0	0	0	0
soote	1	0	0	0	0	0	...	0	0	0	0	0	1
droghte	2	0	0	1	0	0	...	0	0	0	0	0	0
march	1	0	0	0	0	0	...	0	0	0	0	0	0
perced	1	0	0	0	0	0	...	0	0	0	0	0	0
roote	2	0	0	1	0	0	...	0	3	0	0	0	9
bathed	1	1	0	0	0	0	...	0	0	0	0	0	0
veyne	1	2	0	0	0	0	...	0	1	0	0	0	4
licour	1	0	0	0	0	0	...	0	0	0	0	0	0
vertu	2	5	0	0	0	0	...	0	0	0	2	0	23
engendred	2	2	0	0	0	0	...	0	0	0	0	0	1
flour	2	4	0	0	0	4	...	0	0	0	1	0	1
zephyrus	1	0	0	0	0	0	...	0	0	0	0	0	0
sweete	2	3	0	12	0	1	...	0	1	1	1	0	4
breeth	1	2	0	0	0	0	...	0	0	2	0	0	0
inspired	1	0	0	0	0	0	...	0	0	0	0	0	0

Table 3.2 Character-terms matrix from KnT

Rank	arcite	emelye	knyght	palamon	theseus	ypolita
1	palamon 0.875056	howleth 0.776135	unarmed 0.730932	arcite 0.875056	homcomynge 0.942932	femenye 0.915396
2	fightyng 0.810342	ypolita 0.764318	wilnen 0.710335	breme 0.805577	perturben 0.942932	scithia 0.915396
3	breme 0.786654	palamon 0.701573	walweth 0.656901	vale 0.78538	duc 0.901164	weddede 0.843446
4	vale 0.78576	flotery 0.679278	defaute 0.640753	galgopheye 0.78538	perotheus 0.796162	ladys 0.841977
5	galgopheye 0.78576	ruggy 0.679278	machabeus 0.637651	howleth 0.765481	howleth 0.783477	amazones 0.809468
6	foughten 0.77807	asshy 0.679278	wilnest 0.626357	fightyng 0.761243	gliterynge 0.779864	asseged 0.809468
7	myghtest 0.772363	ydropped 0.679278	pavye 0.609359	foughten 0.749968	atthenes 0.759235	onward 0.771976
8	frothen 0.772363	rewefulleste 0.679278	delyt 0.609359	launde 0.737491	fightyng 0.727425	emelye 0.764318
9	bores 0.750824	crynge 0.672517	appetyt 0.609359	belmarye 0.710071	arcite 0.702727	queene 0.73736
10	palamoun 0.744962	egeus 0.667388	hight 0.606959	myghtest 0.704655	perpetuely 0.698612	wonnen 0.711035

Table 3.3 Terms-terms matrix from the whole of the *Tales*

Rank	chivalrie	curteisie	fredom	honour	trouthe	worthy
1	fredom 0.725016	worthy 0.603513	knyghthede 0.781358	processioun 0.708838	forthynketh 0.75342	vavasour 0.967212
2	worthy 0.698166	fredom 0.583385	chivalrie 0.743523	carien 0.708838	grantmercy 0.68172	huberd 0.885131
3	vavasour 0.652694	vavasour 0.562479	honour 0.677404	abbay 0.708838	sworne 0.68172	padowe 0.744168
4	petro 0.630428	padowe 0.53289	bachilrie 0.675938	fredom 0.677404	embassadours 0.66481	moreover 0.741883
5	huberd 0.62671	chivalrie 0.52992	servaunt 0.66177	europe 0.669379	whiderward 0.64917	thralle 0.741883
6	liges 0.618513	avalen 0.521108	non 0.599475	lyle 0.640891	cherlyssh 0.64917	chivalrie 0.658617
7	rit 0.598401	pilates 0.521108	worthy 0.594695	whe 0.640891	trede 0.644798	privilege 0.649963
8	charles 0.586281	huberd 0.520236	curteisie 0.583385	reverenced 0.636312	compelle 0.630573	mysuseth 0.649963
9	olyver 0.586281	loved 0.499779	flour 0.565417	custaunce 0.633946	sweri yng 0.630573	defaute 0.625233
10	armorike 0.586281	knight 0.492656	vavasour 0.539609	chivalrie 0.623794	condicions 0.630573	chaungeable 0.618533

Table 3.4 Characters-terms matrix from WBT

Rank	yonge	wife	wommen	trewe	queene	gentil	worthy
1	meeke 0.995746	fulfild 0.951448	moost 0.968718	aventure 0.990709	mayde 0.961006	folwen 0.956362	contraried 0.963526
2	parfit 0.995746	land 0.951448	specially 0.960957	elles 0.990709	kyng 0.899638	vertuous 0.951425	wyf 0.96237
3	fressh 0.995746	britons 0.951448	breſt 0.960957	tweye 0.990709	chese 0.885707	nys 0.931163	wydwe 0.956866
4	wyves 0.995746	fayerye 0.951448	goost 0.960957	humble 0.990709	yourselven 0.837407	noble 0.926916	devyse 0.946781
5	sende 0.995746	hevene 0.93861	sorweſul 0.960957	deye 0.990709	liſt 0.83456	sooth 0.921278	bar 0.946781
6	diſpence 0.995746	underſtonde 0.93861	ſaufly 0.95929	diſpleſe 0.990709	wheither 0.825526	trewely 0.921278	fouler 0.946781
7	peſtilence 0.995746	mayden 0.93861	deſire 0.942803	repair 0.990709	paraventure 0.824873	clawe 0.921278	heer 0.944797
8	houſbondes 0.995746	jheſus 0.93861	nam 0.942803	cauſe 0.98864	wille 0.822331	galle 0.921278	ſovereyn 0.944739
9	jheſu 0.995746	dayes 0.934591	mooder 0.942803	take 0.980128	arthour 0.821522	kike 0.921278	mercy 0.944739
10	ſhorte 0.995746	bath 0.91344	withinne 0.930809	yong 0.97615	arthures 0.820542	ſeen 0.917429	ſtirte 0.944739

Table 3.5 Characters-terms matrix from SqT

Rank	manly	sire	lady	lordes	corage	list	devout
1	bette 0.997111	ryden 0.993633	hoote 0.901067	knyght 0.962384	therto 0.997263	ride 0.964469	parys 0.992767
2	seide 0.997111	bitwix 0.993633	face 0.901067	halle 0.958118	yong 0.997263	doone 0.964469	yeven 0.992767
3	message 0.997111	leyde 0.993633	colerik 0.901067	kyng 0.947754	bachelor 0.997263	place 0.932362	fully 0.992767
4	leere 0.997111	trippe 0.993633	cleer 0.901067	thridde 0.944706	stable 0.997263	pyn 0.903866	swoor 0.992767
5	vice 0.997111	reyne 0.993633	mansioun 0.901067	nobleye 0.944706	riche 0.997263	ryde 0.893699	knees 0.992767
6	lettre 0.997111	mooten 0.993633	martes 0.901067	deliciously 0.944706	hardy 0.997263	contree 0.893699	lovere 0.992767
7	used 0.997111	seyne 0.993633	joly 0.901067	herknyng 0.944706	sworn 0.997263	nempne 0.893699	tigre 0.992767
8	silable 0.997111	anywhere 0.993633	exaltacioun 0.901067	mynstralles 0.944706	honorable 0.997263	effect 0.878746	humblesse 0.992767
9	accordant 0.997111	ere 0.991118	aries 0.901067	glas 0.928309	yliche 0.997263	abyde 0.878746	troye 0.992767
10	techeth 0.997111	stant 0.982024	canacee 0.881112	amende 0.92745	just 0.997263	descende 0.878746	jason 0.992767

CHAPTER 4

CONCLUSIONS

Now that we have discussed the process of creating and conditioning the data sets, we can turn our attention to some specific calculations' results and their interpretations. In previous chapters, we developed a sort of list of necessary characteristics of a chivalric knight from an examination of Chaucer's own description of the Knight and the existing scholarship on the topic. As an exemplar of the heroic knight, Chaucer's is "worthy". He values truth, honor and chivalry above all other ideals. He is humble, unconcerned with the accumulation of treasure. He is brave and certain in his actions because he knows that God will take care of him.

These are the virtues that we can presume Chaucer intends (in the lightest sense of the word) for his audience to associate with the Knight. By examining these characteristics through this semantic model, we can determine more quantitatively if those characteristics apply to any of the knightly characters at all. Throughout this chapter, we will be referring to a series of tables (Tables 4.1-4.12) that show the most significant semantic relationships between the primary characters in "The Knight's Tale" (Arcite, Palamon, Emelye, Theseus, Ypolita) and some of the things we might associate with a chivalrous character ("knyght" itself, "chivalrie" itself, "curteisie", "honour", "trouthe", "worthy"). From here we can begin to answer some questions that a traditional analysis of *The Canterbury Tales* might leave unanswered. What are we to make of the

pilgrim Knight as a representation of the real-world and other literary chivalric knights that Chaucer's audience would be familiar with?

“The Knight's Tale” offers us the opportunity to, at least temporarily, avoid the ambiguity of the word “knyght” by examining some of the named characters in the tale itself. Beginning with Arcite we see, as we might expect, that he is highly correlated with his brother-knight Palamon (both as “Palamon” and as “Palamoun”) and combat or physical prowess in a number of forms (“fightyng”, “foughten”, “myghtest”). We can see the high similarity score between them (0.875056) directly and reflected by the surface-level similarity of their lists of associated terms (Table 3.2). In terms of semantically significant sentences associated with the term “Arcite”, we again see high correlations with both “Palamon” and combat (Table 4.1). The same examination, this time performed on “Emelye” will show some major differences in the results.

Emelye is most closely associated with her mother Ypolita, the knight Palamon, and the negative emotions associated with “howleth” and “cryngge”. As we might expect from Arcite's list, but contrary to what we would expect given Palamon's, “Emelye” is positively correlated with one of the knights (Table 3.2). This is reinforced by the sentence matrix found in Table 4.2. The ten most-strongly correlated sentences for “Emelye” describe her final wedding to Palamon after Arcite's death, her life prior to the arrival of the knights, and her relationships with Theseus and Ypolita, among other things. We should note, however, that Arcite never appears directly in this matrix, and the only sentence indirectly relating to him occurs after his death.

Unlike his friend Arcite, at least one of our measures shows a significant correlation between Palamon and Emelye – our terms-sentences matrix (Table 4.4).

Palamon's second-ranked sentence is identical to Emelye's first-ranked, and Emelye is mentioned again in an only slightly-less correlated sentence (rank eight). Still, Palamon's sentence matrix is naturally replete with references to Arcite. A limitation of the scope of this study results in some probably meaningless correlations appearing. "Vale" and "Galgopheye" appear only together, and given the small sample size inherent in such a narrow focus as a single work, there is simply not enough data to force some of the more trivial data like this out of view. Similarly, "wel fynden that Arcite and Palamo[u]n" is hardly a meaning-rich sentence in itself, despite the semantic similarity between it and the knights.

Examining Table 3.3, some of the terms we found to be generally related to the Knight and knighthood reveal interesting complications. The semantic similarity matrix for "chivalrie" contains no instances of "Arcite", "Palamon", "knyght" or any of the other characters we've discussed. "Freedom", "worthy" and "honour" prominently appear, as expected. Indeed, we can see that their columns share many of the same terms. We can infer from the score of the tenth-ranked term that even if one of our characters were to appear, their score would indicate too low a correlation to be of interest. "Trouthe", however, is far removed from the rest, sharing no major terms with the others and with only "knyght" appearing in its sentence matrix (and that coming low in the rankings). There is perhaps no more fitting outlier, given the question of the true nature of the Knight.

Incidentally, Huberd the Friar does appear in the "trouthe" column with a strong correlation score. Other Canterbury pilgrims seem to be more semantically related to those virtues that we are supposed to associate with the Knight. If we look to Table 4.7,

though, we see that the second-ranked sentence is the first line of the Knight's portrait in the GP. While perhaps the concept of "chivalrie" is related to knighthood, it seems the relationship is not reciprocal. Again, we see a limitation of the sample size in the repetition of some higher-profile sentences because they include so many of the terms we're searching for. The fourth-ranked sentence for "freedom", for example, includes "chivalrie", and several include the words "honour" and "trouthe". Such prominent and repeated but superficial correlations may obscure some more relevant data points, regardless of the precautions taken to avoid it.

So, what are we to make of Chaucer's Knight? The correlation calculations indicate that none of the expected terms are closely associated with the Knight. The term "knyght" itself, most closely correlated with "unarmed" in Table 3.2 hardly seems to resemble the champion of virtue that Chaucer described. Those terms, in contrast, are associated with the characters of "The Knight's Tale" rather than the Knight himself. And finally, as we've noted, there appears to be no connection between the Knight's characterization and his tale's content. Thus, any indirect connection we might make is disqualified on the face of it. The Knight (and the concept of knighthood, by extension) is far removed from the fanciful ideals of a chivalric hero. Instead, Palamon and Arcite appear to fulfil the role of the "proper" (or at least "expected") chivalric hero.

This is yet another manifestation of the dichotomy between the chivalric ideal and the realities of medieval life. The Knight is somehow both idealized and realistic in his presentation. He's described as a nearly perfect specimen of knighthood, yet the underlying semantic content of the *Tales* seems to give Chaucer the lie. However, he also propagates the chivalric ideal through his tale. Palamon and Arcite, even with all their

humanity, represent the virtues that knighthood is supposed to entail. In that sense, we can say that Chaucer's Knight *is* the stereotypical knight of the day. The Knight embodies the difference between the reality of a chivalric hero and the literary ideal to which he is supposed to adhere. Of course, he fails to live up to that ideal. This is hardly a surprise - as we've seen, were he a real person, no one would expect otherwise of him. Still, he appears to represent the worldview that Johan Huizinga and others described, putting chivalry at the core of his existence.

That the Knight does not revel in violence, even when one might expect that he has cause to do so, is perhaps simply a consequence of the nature of the *Tales*. It would be an ill-fated journey that included the possibility of violence against one pilgrim by another. As Frieda Penninger noted, though, the Knight does not shy away from gruesome descriptions of death, and he appears to show no interest in the more spiritual matters that are important to some of the other characters. We can take from this then, that the Knight is as complicated a character as any real-world knight of the day. Neither wholly violent and brutish, nor possessed of the pure virtue of the ideal knight.

On the other hand, we can see in the contrast between the Knight and the Squire something of the differences that Nigel Saul pointed out between the two opposing views of chivalry: that either the knights were too violent, or not nearly violent enough. The Knight, representing the previous group, has proven himself in many battles. He is not especially well-read, and not particularly inclined to infuse his tale with his military experience. His description of Arcite's death is perhaps simply an attempt to impress the other pilgrims with his knowledge of the results of war. Finally, while occupying the

highest social status among all the pilgrims, he is described as being outwardly rather modest, and not overly concerned with intellectual pursuits.

By contrast, his son the Squire seems to represent the Knight's complement. Less experienced in battle, the Squire is described as young, obsessed with love, and well-clothed. His tale, though interrupted, is more epic in scale compared to the somewhat constrained subject matter of his father's. Critically, it fails to reach the heroic action that he promises. Instead, it focuses on a series of magical artifacts and a subsidiary tale of lost love. In nearly all senses, he is the epitome of the well-rounded chivalric hero. He combines both the learning and some of the prowess that we have come to expect from such a man, while being a paragon of neither.

An expansion of this study to include a broader corpus of medieval works would go a long way to alleviating the sample size problems we've observed here. The specific methods would have to be tailored to each work depending on its encoding scheme, and we would need to develop a scheme to reconcile variations of spelling across a multitude of authors and editions.

These concepts of 14th Century European chivalry, of course, appear in other works from other traditions. That is, it is hardly only English knights that are described in such glowing terms. It is perhaps less common to find a knightly character telling a tale of knightly virtue, so the differences between Chaucer's Knight's presentation and the knights in his tale will be unlikely to surface again. The additional data an analysis of those works would provide would help improve the accuracy of the various calculations employed herein.

Table 4.1 Sentences matrix for “Arcite”

Rank	Score	Sentence
1	0.966874	wel fynden that arcite and palamoun
2	0.866909	and right so ferden they with palamon
3	0.803831	and in the grove, at tyme and place yset, this arcite and this palamon ben met
4	0.801101	and whan this duc was come unto the launde, under the sonne he looketh, and anon he was war of arcite and palamon, that foughten breme, as it were bores two
5	0.797878	but stynte i wole of theseus a lite, and speke of palamon and of arcite
6	0.796446	thou myghtest wene that this palamon in his fightyng were a wood leon, and as a cruuel tigre was arcite; as wilde bores gonne they to smyte, that frothen whit as foom for ire wood
7	0.788775	ther nas no tygre in the vale of galgopheye, whan that hir whelp is stole whan it is lite, so cruuel on the hunte as is arcite for jelous herte upon this palamon
8	0.776776	men seyde eek that arcite shal nat dye; he shal been heeled of his maladye
9	0.758003	ne in belmarye ther nys so fel leon, that hunted is, or for his hunger wood, ne of his praye desireth so the blood, as palamon to sleen his foo arcite
10	0.734745	now wol i stynte of palamon a lite, and lete hym in his prisoun stille dwelle, and of arcita forth i wol yow telle

Table 4.2 Sentences matrix for “Emelye”

Rank	Score	Sentence
1	0.855502	and thus with alle blisse and melodye hath palamon ywedded emelye
2	0.799919	shrighte emelye, and howleth palamon, and theseus his suster took anon swownynge, and baar hire fro the
3	0.77901	cleer was the day, as i have toold er this, and theseus with alle joye and blis, with his ypolita, the faire queene, and emelye, clothed al in grene, on huntyng be they riden roially
4	0.772305	thus endeth palamon and emelye; and god save al this faire compaignye! amen
5	0.764119	tho sente theseus for emelye
6	0.723052	upon the right hond wente olde egeus, and on that oother syde duc theseus, with vessels in hir hand of gold ful fyn, al ful of hony, milk, and blood, and wyn; eek palamon, with ful greet compaignye; and after that cam woful emelye, with fyr in honde, as was that tyme the gyse, to do the office of funeral servyse
7	0.70274	tho cam this woful theban palamoun, with flotery berd and ruggy, asschy heeres, in clothes blake, ydropped al with teeres; and, passynge othere of wepynge, emelye, the rewefulleste of al the compaignye
8	0.702685	i am thy mortal foo, and it am i that loveth so hoote emelye the brighte that i wol dye present in hir sighte
9	0.676816	and right so ferden they with palamon
10	0.673375	it nas nat of the day yet fully pryme whan set was theseus ful riche and hye, ypolita the queene, and emelye, and othere ladys in degrees aboute

Table 4.3 Sentences matrix for “knyght”

Rank	Score	Sentence
1	0.927402	wo was this knyght, and sorwefully he siketh; but what! he may nat do al as hym liketh
2	0.88565	and whan this knyght hath thus his tale toold, he rideth out of halle, and doun he lighte
3	0.804809	and to the thridde knyght right thus he seith, thou hast nat doon that i comanded thee
4	0.767526	but fynally the kyng axeth this knyght the vertu of this courser and the myght, and preyde hym to telle his governaunce
5	0.758938	this knyght is to his chambre lad anoon, and is unarmed, and to mete yset
6	0.745165	this false knyght, that hath this tresoun wroght, berth hire on hond that she hath doon thys thyng
7	0.736635	for if ther fille tomorwe swich a cas, ye knowen wel that every lusty knyght that loveth paramours and hath his myght, were it in engelond or elleswhere, they wolde, hir thankes, wilnen to be there, to fighte for a lady, benedicitee! it were a lusty sighte for to see
8	0.725447	have heer my trouthe, quod the knyght, i grante
9	0.695485	he may nat tempte yow over youre myght, for crist wol be youre champion and knyght
10	0.689835	he was a verray, parfit gentil knyght

Table 4.4 Sentences matrix for “Palamon”

Rank	Score	Sentence
1	0.986105	and right so ferden they with palamon
2	0.845992	and thus with alle blisse and melodye hath palamon ywedded emelye
3	0.83452	this palamon gan knytte his browes tweye
4	0.834397	wel fynden that arcite and palamoun
5	0.8122	and whan this duc was come unto the launde, under the sonne he looketh, and anon he was war of arcite and palamon, that foughten breme, as it were bores two
6	0.810262	but stynte i wole of theseus a lite, and speke of palamon and of arcite
7	0.802443	and in the grove, at tyme and place yset, this arcite and this palamon ben met
8	0.781665	shrighte emelye, and howleth palamon, and theseus his suster took anon swownynge, and baar hire fro the corps away
9	0.779593	ther nas no tygre in the vale of galgopheye, whan that hir whelp is stole whan it is lite, so crueel on the hunte as is arcite for jelous herte upon this palamon
10	0.778873	now wol i stynte of palamon a lite, and lete hym in his prisoun stille dwelle, and of arcita forth i wol yow telle

Table 4.5 Sentences matrix for “Theseus”

Rank	Score	Sentence
1	0.958882	what fold been ye, that at myn homcomynge perturben so my feste with crynge? quod theseus
2	0.819591	duc theseus, with al his compaignye, is comen hoom to atthenes his citee, with alle blisse and greet solempnitee
3	0.816746	and in this wise i lete hem fightyng dwelle, and forth i wole of theseus yow telle
4	0.806252	tho sente theseus for emelye
5	0.798104	shrighte emelye, and howleth palamon, and theseus his suster took anon swownyng, and baar hire fro the corps away

Table 4.6 Sentences matrix for “Ypolita”

Rank	Score	Sentence
1	0.851025	cleer was the day, as i have toold er this, and theseus with alle joye and blis, with his ypolita, the faire queene, and emelye, clothed al in grene, on huntyng be they riden roially
2	0.847828	ful many a riche contree hadde he wonne; what with his wysdom and his chivalrie, he conquered al the regne of femenye, that whilom was ycleped scithia, and weddede the queene ypolita, and broghte hire hoom with hym in his contree with muchel glorie and greet solempnytee, and eek hir yonge suster emelye
3	0.839852	it nas nat of the day yet fully pryme whan set was theseus ful riche and hye, ypolita the queene, and emelye, and othere ladys in degrees aboute
4	0.806143	and certes, if it nere to long to heere, i wolde have toold yow fully the manere how wonnen was the regne of femenye by theseus and by his chivalrye; and of the grete bataille for the nones bitwixen atthenes and amazones; [page 26] and how asseged was ypolita, the faire, hardy queene of scithia; and of the feste that was at hir weddyng, and of the tempest at hir hoom-comyng; but al that thyng i moot as now forbere
5	0.757525	no neer atthenes wolde he go ne ride, ne take his ese fully half a day, but onward on his wey that nyght he lay, and sente anon ypolita the queene, and emelye, hir yonge suster sheene, unto the toun of atthenes to dwelle, and forth he rit; ther is namoore to telle

Table 4.7 Sentences matrix for “chivalrie”

Rank	Score	Sentence
1	0.777789	arveragus, with heele and greet honour, as he that was of chivalrie the flour, is comen hoom, and othere worthy men
2	0.705954	the knight's portrait a knyght ther was, and that a worthy man, that fro the tyme that he first bigan to riden out, he loved chivalrie, trouthe and honour, fredom and curteisie
3	0.634812	was nowher swich a worthy vavasour
4	0.592093	and whan this worthy duc hath thus ydon, he took his hoost, and hoom he rit anon with laurer crowned as a conquerour; and ther he lyveth in joye and in honour terme of his lyf; what nedeth wordes mo? and in a tour, in angwissh and in wo, this palamon and his felawe arcite for everemoore; ther may no gold hem quite
5	0.586433	this worthy lymytour was cleped huberd

Table 4.8 Sentences matrix for “curteisie”

Rank	Score	Sentence
1	0.716247	the knight's portrait a knyght ther was, and that a worthy man, that fro the tyme that he first bigan to riden out, he loved chivalrie, trouthe and honour, fredom and curteisie
2	0.576034	the noble usage of freres yet is this, the worthy men of hem shul first be served; and certainly he hath it well disserved
3	0.571987	was nowher swich a worthy vavasour
4	0.551288	i wol yow telle a tale which that i lerned at padowe of a worthy clerk, as preved by his wordes and his werk
5	0.54132	this parissch clerk, this joly absolon, hath in his herte swich a love-longynge that of no wyf took he noon offrynge; for curteisie, he seyde, he wolde noon

Table 4.9 Sentences matrix for “freedom”

Rank	Score	Sentence
1	0.814181	the knight's portrait a knyght ther was, and that a worthy man, that fro the tyme that he first bigan to riden out, he loved chivalrie, trouthe and honour, fredom and curteisie
2	0.778905	and juppiter so wys my soule gye, to speken of a servaunt proprely, with alle circumstanes trewely that is to seyen, trouthe, honour, knyghthede, wysdom, humblesse, estaat, and heigh kynrede, fredom, and al that longeth to that art so juppiter have of my soule part, as in this world right now ne knowe i non so worthy to ben loved as palamon, that serveth yow, and wol doon al his lyf
3	0.691841	he was of knyghthod and of fredom flour; fortune hym made the heir of hire honour
4	0.679926	he was therwith fulfild of gentillesse, of honour, and of parfit worthynesse
5	0.67821	this phebus, that was flour of bachilrie, as wel in fredom as in chivalrie, for his desport, in signe eek of victorie of phitoun, so as telleth us the storie, was wont to beren in his hand a bowe
6	0.666427	arveragus, with heele and greet honour, as he that was of chivalrie the flour, is comen hoom, and othere worthy men
7	0.560573	was nowher swich a worthy vavasour
8	0.537529	this worthy lymytour was cleped huberd
9	0.521091	a, quod melibee, now se i wel that ye loven nat myn honour ne my worshipe.
10	0.510871	this maked emelye have remembraunce to doon honour to may, and for to ryse

Table 4.10 Sentences matrix for “honour”

Rank	Score	Sentence
1	0.834412	he was therwith fulfild of gentillesse, of honour, and of parfit worthynesse
2	0.775376	a, quod melibee, now se i wel that ye loven nat myn honour ne my worshipe.
3	0.725623	this maked emelye have remembraunce to doon honour to may, and for to ryse
4	0.719247	this child with pitous lamentacioun up taken was, syngynge his song alway, and with honour of greet processioun they carien hym unto the nexte abbay
5	0.703705	arveragus, with heele and greet honour, as he that was of chivalrie the flour, is comen hoom, and othere worthy men
6	0.682554	thou ne hast nat doon to hym swich honour and reverence as thee oughte,
7	0.659467	wherfore in laude, as i best kan or may, of thee and of the white lyle flour which that the bar, and is a mayde alway, to telle a storie i wol do my labour; nat that i may encressen hir honour, for whe hirself is honour and the roote of bountee, next hir sone, and soules boote
8	0.656681	irreverence is whan men do nat honour there as hem oghte to doon, and waiten to be revered.
9	0.656649	i prey to god in honour hire susteene, and wolde she were of al europe the queene
10	0.6552	he was of knyghthod and of fredom flour; fortune hym made the heir of hire honour

Table 4.11 Sentences matrix for “trouthe”

Rank	Score	Sentence
1	0.981039	it stryveth eek alday agayn trouthe
2	0.954086	but now at erst in trouthe oure dwellyng is
3	0.8827	she seyde she was so mazed in the see that she forgat hir mynde, by hir trouthe
4	0.877464	i graunte, quod that oother, out of doute, that, by my trouthe, i wol thee nat biwreye
5	0.840101	my trouthe wol i kepe, i wol nat lye
6	0.831323	but by my trouthe, if thou were seculer, thou woldest ben a trede-foul aright
7	0.792446	thou shalt sweren eek in doom, whan thou art constreyned by thy domesman to witnessen the trouthe.
8	0.757307	and she gan kisse his brest, that herde this, and was ful glad he koude trouthe espye
9	0.757286	that me forthynketh, quod this januarie, he is a gentil squier, by my trouthe! if that he deyde, it were harm and routhe
10	0.729271	have heer my trouthe, quod the knyght, i grante

Table 4.12 Sentences matrix for “worthy”

Rank	Score	Sentence
1	0.981143	was nowher swich a worthy vavasour
2	0.91757	thus starf this worthy, myghty hercules
3	0.902457	this worthy lymytour was cleped huberd
4	0.783424	for sothe he was a worthy man with alle, but, sooth to seyn, i noot how men hym calle
5	0.768681	i wol yow telle a tale which that i lerned at padowe of a worthy clerk, as preved by his wordes and his werk
6	0.760551	and yet seith this pamphilles moreover that they that been thralle and bonde of lynage shullen been maad worthy and noble by the riches
7	0.749338	the miller's prologue whan that the knyght had thus his tale ytoold, in al the route nas ther yong ne oold that he ne seyde it was a noble storie, and worthy for to drawn to memorie; and namely the gentils everichon
8	0.745812	the knight's portrait a knyght ther was, and that a worthy man, that fro the tyme that he first bigan to riden out, he loved chivalrie, trouthe and honour, fredom and curteisie
9	0.693754	this worthy duc answerde anon agayn, and seyde, this is a short conclusioun
10	0.692174	for job seith, synful men doon werkes worthy of confusioun.

WORKS CITED

- Anderson, David. "The Fourth Temple of The Knight's Tale: Athenian Clemency."
Studies in the Age of Chaucer 2 (1986): 113-125.
- Arnold, Taylor and Lauren Tilton. *Humanities Data in R Exploring Networks, Geospatial Data, Images, and Text*. Springer International Publishing, 2016.
- Barber, Richard. *The Reign of Chivalry*. Woodbridge: The Boydell Press, 2005.
- Barr, Helen. "Chaucer's Knight: A Christian Killer?" *The English Review* 12.2 (2001).
- Bartholomew, Barbara. *Fortuna and Natura: A Reading of Three Chaucer Narratives*. De Gruyter Mouton, 2015.
- Bellegarda, Jerome R. "Latent Semantic Mapping." *IEEE Signal Processing Magazine* September 2005: 70-80.
- Benson, Larry D. and John Leyerle, *Chivalric Literature: Essays on Relations Between Literature & Life in the Later Middle Ages*. Kalamazoo: Medieval Institute Publications, 1980.
- Benson, Larry D. *Contradictions: From Beowulf to Chaucer*. Ed. Theodore M. Andersson and Stephen A. Barney. Aldershot: Ashgate Publishing Company, 1995.

- Benson, Robert G. and Susan J. Ridyard, *New Readings of Chaucer's Poetry*. Suffolk: St Edmundsbury Press, Ltd, 2003.
- Blanch, Robert J. and Julian N. Wasserman. *From Pearl to Gawain Forme to Fynisment*. Gainesville: University Press of Florida, 1995.
- Bode, Katherine. "Literary Studies in the Digital Age." Bode, Katherine. *Reading by Numbers*. Anthem Press, 2012.
- . "The Equivalence of 'Close' and 'Distant' Reading; or, Toward a New Object for Data-Rich Literary History." *Modern Language Quarterly* 78.1 (2017): 77-106.
- Boitani, Peter and Anna Torti, *Literature in Fourteenth-Century England*. Cambridge: D. S. Brewer, 1983.
- Bouma, Gerlof. "Normalized (Pointwise) Mutual Information in Collocation Extraction." 2009.
- Brewer, Derek S. "Chaucer's Knight as Hero, and Machaut's Prise d'Alexandrie." *Heroes and Heroines in Medieval English Literature*. Ed. Leo Carruthers. Cambridge: D. S. Brewer, 1994. 81-96.
- . "Class Distinction in Chaucer." *Speculum* 43.2 (1968): 290-305.
- Bruggerman, Jeroen. "Methods: Data and Software & Glossary of Network Notions." *Social Networks: An Introduction*. 2008. 110-139.
- Bumke, Joachim. *The Concept of Knighthood in the Middle Ages*. Trans. W. T. H. Jackson and Erika Jackson. New York: AMS Press, 1982.

- Burnley, David. *Courtliness and Literature in Medieval England*. New York: Addison Wesley Longman Inc., 1998.
- Chance, Jane. "Representing Rebellion: The Ending of Chaucer's Knight's Tale and the Castration of Saturn." *Studia Anglica Posnaniensia* 38 (2002).
- Chaucer, Geoffrey. "[OTA] The Canterbury Tales." 1957. *Oxford Text Archive*. Ed. Fred Norris Robinson. <<http://purl.ox.ac.uk/ota/3228>>.
- . *The Riverside Chaucer*. Ed. Larry D. Benson. Oxford University Press, n.d.
- Clark, Stephen. "Vector Space Models of Lexical Meaning." *The Handbook of Contemporary Semantic Theory*. Ed. Shalom Lappin and Chris Fox. John Wiley & Sons, Inc., 2015. 493-520.
- Clasby, Eugene. "Chaucer's Constance: Womanly Virtue and the Heroic Life." *The Chaucer Review* 13.3 (1979): 221-233.
- Clein, Wendy. *Concepts of Chivalry in Sir Gawain and the Green Knight*. Norman, Oklahoma: Pilgrim Books, 1987.
- Cole, Andrew. *Literature and Heresy in the Age of Chaucer*. Cambridge: Cambridge University Press, 2008.
- Corrigan, Nora. "The Knight's Earnest Game in Chaucer's The Canterbury Tales." *Games and Gaming in Medieval Literature*. Ed. Serina Patterson. Palgrave MacMillan, 2015.

Curtis, Carl C. III. "Biblical Analogy and Secondary Allegory in Chaucer's The Knight's Tale." *Christianity and Literature* 57.2 (2007).

Daniel, Norman. *Heroes and Saracens: An Interpretation of the Chansons de Geste*.
Edinburgh: Edinburgh University Press, 1984.

Davis, Isabel. *Writing Masculinity in the Later Middle Ages*. Cambridge University Press, 2007.

Deerwester, Scott, et al. "Indexing by Latent Semantic Analysis." *Journal of the American Society for Information Science* 41.6 (1990): 391-407.

Drucker, Johanna. "Why Distant Reading Isn't." *PMLA* 132.3 (2017): 628-635.

Everett, Dorothy. *Essays on Middle English Literature*. Ed. Patricia Kean. Oxford University Press, 1959.

Finnegan, Robert Emmett. "A Curious Condition of Being: the City and the Grove in Chaucer's Knight's Tale." *Studies in Philology* (2009).

Fleming, Paul. "Tragedy, for Example: Distant Reading and Exemplary Reading (Moretti)." *New Literary History* 48.3 (2017): 437-455.

Gellrich, Jesse M. *Discourse and Dominion in the Fourteenth Century: Oral Contexts of Writing in Philosophy, Politics, and Poetry*. Princeton University Press, 1995.

Gries, Stefan Thomas. *Statistics for Linguistics with R: A Practical Introduction*. De Gruyter Mouton, 2009.

- Guidry, Marc S. "The Parliaments of Gods and Men in The Knight's Tale." *The Chaucer Review* 43.2 (2008).
- Hartigan, J. A. and M. A. Wong. "Algorithm AS 136: A K-Means Clustering Algorithm." *Journal of the Royal Statistical Society, Series C (Applied Statistics)* 28.1 (1979): 100-108.
- Huizinga, Johan. *The Waning of the Middle Ages*. Garden City: Doubleday Anchor Books, 1954.
- Huppe, Bernard F. "The Concept of the Hero in the Early Middle Ages." *Concepts of the Hero in the Middle Ages and the Renaissance*. Ed. Norman T. Burns and Christopher J. Reagan. Albany: State University of New York, 1975.
- Jin, Jay. "Problems of Scale in "Close" and "Distant" Reading." *Philological Quarterly* 96.1 (2017): 105-129.
- Jockers, Matthew L. *Macroanalysis: Digital Methods & Literary History*. University of Illinois Press, 2017.
- . *Text Analysis with R for Students of Literature*. Springer, 2014.
- Jones, Lindsey M. "Chaucer's Anxiety of Poetic Craft: The Squire's Tale." *Style* 41.3 (2007).
- Jones, Terry. *Chaucer's Knight: The portrait of a medieval mercenary*. London: Weidenfeld and Nicolson, 1980.

- Jucker, Andreas H. "Courtesy and Politeness in Sir Gawain and the Green Knight." *Studia Anglica Posnaniensia* 49.3 (2014).
- Kaeuper, Richard W. *Chivalry and Violence in Medieval Europe*. Oxford: Oxford University Press, 1999.
- Keen, Maurice. *Chivalry*. New Haven: Yale University Press, 1984.
- Landauer, Thomas K. "LSA as a Theory of Meaning." *Handbook of Latent Semantic Analysis*. Ed. Thomas K. Landauer, et al. Lawrence Erlbaum Associates, 2007. 3-33.
- Long, Hoyt and Richard Jean So. "Turbulent Flow." *Modern Language Quarterly* 77.3 (2016): 345-367.
- Malone, Kemp. *Chapters on Chaucer*. Baltimor: The Johns Hopkins Press, 1951.
- Manning, Christopher D. *Introduction to Information Retrieval*. Cambridge University Press, 2008.
- Martin, Dian I. and Michael W. Berry. "Mathematical Foundations Behind Latent Semantic Analysis." *Handbook of Latent Semantic Analysis*. Ed. Thomas K. Landauer, et al. London: Lawrence Erlbaum Associates, 2007.
- Mayrhofer, Sonja. "'This sely jalous housbonde to bigyle': Reading and Performance in Chaucer's The Miller's Tale." *Philological Quarterly* 97.4 (2018).
- McCormack, Frances. *Chaucer and the Culture of Dissent: The Lollard context and subtext of the Parson's Tale*. Chippenham: Antony Rowe Ltd, 2007.

Mills, Charles. *The History of Chivalry; or Knighthood and Its Times*. London: Longman, Rees, Ormes, Brown and Green, 1826.

Minnis, Alastair J. and Stephen H. Rigby, *Historians on Chaucer: The "General Prologue" to the Canterbury Tales*. Oxford: Oxford University Press, 2014.

Moretti, Franco. "A Response." *PMLA* 132.3 (2017): 686-689.

—. "Conjectures on World Literature." *New Left Review* 1 (2014): 54-68.

—. *Graphs, Maps, Trees*. Verso, 2005.

—. "Network Theory, Plot Analysis." *Pamphlets of the Stanford Literary Lab* (2011).

—. "Style, Inc. Reflections on Seven Thousand Titles (British Novels, 1740-1850)." *Critical Inquiry* 36.1 (2009): 134-158.

Morgan, Gerald. "The Worthiness of Chaucer's Worthy Knight." *The Chaucer Review* 44.2 (2009).

Mueller, Derek Norton. "Clouds, Graphs, and Maps: Distant Reading and Disciplinary Imagination." PhD Dissertation. 2009.

Muscatine, Charles. *Medieval Literature, Style, and Culture*. Columbia: University of South Carolina Press, 1999.

Nyhan, Julianne. "Text Encoding and Scholarly Digital Editions." Warwick, Claire and Melissa M. Terras. *Digital Humanities in Practice*. Facet Publishing, 2012.

O'Brien, Timothy D. "'Sikernesse" and "Fere" in "Troilus and Criseyde"." *The Chaucer Review* 38.3 (2004): 276-293.

Penninger, Frieda Elaine. *Chaucer's Troilus and Criseyde and The Knight's Tale: Fictions Used*. Lanham: University Press of America, 1993.

Powicke, Michael. *Military Obligation in Medieval England*. Oxford: Clarendon Press, 1962.

Pratt, John H. "Was Chaucer's Knight Really a Mercenary?" *The Chaucer Review* 22.1 (1987): 8-27.

Ramsey, Lee C. *Chivalric Romances: Popular Literature in Medieval England*. Bloomington: Indiana University Press, 1983.

Rigby, Stephen H. *Wisdom and Chivalry: Chaucer's Knight's Tale and Medieval Political Theory*. Boston: Brill, 2013.

Robertson, D. W. *A Preface to Chaucer: Studies in Medieval Perspectives*. Princeton: Princeton University Press, 1963.

Rogerson, Margaret. "Reading Chaucer 'In Parts': The Knight's Tale and The Two Noble Kinsmen." *Storytelling: Critical and Creative Approaches*. Ed. Jan Shaw, Philippa Kelly and L. E. Semler. 2013.

Rose, Donald M., ed. *New Perspectives in Chaucer Criticism*. Norman, Oklahoma: Pilgrim Books, Inc., 1981.

- Sanders, Ivor John. *Feudal Military Service in England: A Study of the Constitutional and Military Powers of the Barones in Medieval England*. London: Oxford University Press, 1956.
- Saul, Nigel. *Chivalry in Medieval England*. Cambridge: harvard University Press, 2011.
- Scattergood, John. *Reading the Past: Essays on Medieval and Renaissance Literature*. Portland: Four Courts Press, 1996.
- Shimomura, Sachi. "The Walking Dead in Chaucer's Knight's Tale." *The Chaucer Review* 48.1 (2013): 1-37.
- Steadman, John M. "The Pattern of a Christian Hero." *The Hero in Literature*. New York: Fawcett, 1969. 165-185.
- Stewart, James T. ""They weren no thyng ydel": Noblemen and Their Supporters in Chaucer's Knight's Tale." *The Chaucer Review* 53.3 (2018): 283-307.
- Strohm, Paul. *Social Chaucer*. Harvard University Press, 1994.
- Townsend, Francis G. "Chaucer's Nameless Knight." *The Modern Language Review* 49.1 (1954): 1-4.
- Trumpener, Katie. "Paratext and Genre System: A Response to Franco Moretti." *Critical Inquiry* 36.1 (2009): 159-171.
- Turney, Peter D. and Patrick Pantel. "From Frequency to Meaning: Vector Space Models of Semantics." *Journal of Artificial Intelligence Research* 37 (2010): 141-188.

- Underwood, Ted. "Distant Reading and Recent Intellectual History." Gold, Matthew K. and Lauren F. Klein. *Debates in the Digital Humanities*. University of Minnesota Press, 2016.
- Vale, Malcolm. *War and Chivalry: Warfare and Aristocratic Culture in England, France and Burgundy at the End of the Middle Ages*. Athens: The University of Georgia Press, 1981.
- Wadiak, Walter. "Chaucer's Knight's Tale and the Politics of Distinction." *Philological Quarterly* 89 (2010).
- Weiss, Victoria L. "The Play World and the Real World: Chivalry in Sir Gawain and the Green Knight." *Philological Quarterly* 72.4 (1993).
- Wheatley, Edward. "Murderous Sows in Chaucer's Knight's Tale and Late Fourteenth-Century France." *The Chaucer Review* 44.2 (2009).
- Widdows, Dominic. "Word-Vectors and Search Engines." *Geometry and Meaning*. 2013.
- Wimsatt, James I. "Type Conceptions of the Good Knight in the French Arthurian Cycles, Malory and Chaucer." *Heroes and Heroines in Medieval English Literature*. Ed. Leo Carruthers. Cambridge: D. S. Brewer, 1994. 137-148.

APPENDIX A

R SOURCE

```
##### Canterbury Tales Analysis Phase 1 - Build Database #####
```

```
library(xml2) #For text parsing
library(lsa) #For the basic weighting functions
library(irlba) #For linear algebra stuff
library(Matrix) #For matrix stuff
```

```
##### Function Definitions #####
```

```
ppmi = function(mat) # Normalize with PMI
{
  total = sum(mat, na.rm = T)
  pcol = apply(mat, 2, function(x) sum(x, na.rm = T)/total)
  prow = apply(mat, 1, function(x) sum(x, na.rm = T)/total)
  pmat = as.matrix(prow) %*% t(pcol)
  mat = mat/total
  mat = mat/pmat
  mat = apply(mat, 1, log)
  mat = t(mat)
  mat[mat < 0] = 0
  mat[is.na(mat)] = 0
  return(mat)
}
```

```
similarity = function(mat, vec, fullResults = F, depth=50) # Compute cosine similarity
over a matrix
```

```
{
  cos_sim = function(x, y)
  {
    x %*% y/(sqrt(x %*% x) * sqrt(y %*% y))
  }
  if (class(mat) == "docMatrix")
  {
    mat <- mat@mat
  }
  if (length(vec) == 1)
  {
```

```

keyword = vec
if (keyword %in% row.names(mat) == F)
{
  stop("Your keyword doesn't match any of your matrix's row names.")
}
vec = mat[keyword, ]
}
results = apply(mat, 1, cos_sim, vec)
if (fullResults == F)
{
  results = sort(results, decreasing = T)[1:depth]
}
return(results)
}

semantic_model = function(keyword_vec) # Generate tale-specific semantic model
{
  ord = which(sentence_divs == keyword_vec)
  m_ord = m[,ord]
  mp_ord = ppmi(m_ord)
  ms_ord = irlba(mp_ord, nv = as.integer(sqrt(length(ord))))
  w_or ds = ms_ord$u %*% diag(ms_ord$d)
  rownames(w_or ds) = rownames(m_ord)

  # Build word/topic list for individual tale
  kclusts = kmeans(w_or ds, centers = 50, iter.max=30)
  word_groups = matrix("", 50, 50)
  for (i in 1:nrow(kclusts$centers))
  {
    print(i)
    vec = kclusts$centers[i,]
    sims = similarity(w_or ds, vec)
    word_groups[i,] = paste(names(sims),round(sims,6))
  }
  return(word_groups)
}

#### Starting with actual calculations ####

#Set up some basics, including reading the source HTML file
stop_words_chaucer = c("1", "lyk", "goon", "ay", "thow", "from", "namoore", "wente",
"where", "had", "name", "ay", "under", "se", "cam", "mo", "wey", "she", "hire", "seith",
"wolde", "how", "o", "after", "seyn", "myn", "thise", "fro", "oother", "noght", "til", "up",
"oon", "gan", "shul", "agayn", "koude", "som", "forth", "wise", "go", "doun", "am",
"into", "thyn", "wo", "wole", "see", "shalt", "wordes", "over", "come", "save", "ii", "fitt",
"t", "noot", "bee", "the", "of", "and", "a", "as", "but", "that", "to", "in", "for", "i", "he",

```

```
"is", "ye", "by", "this", "me", "it", "nat", "with", "his", "so", "tale", "my", "yow", "wol",
"may", "now", "al", "telle", "man", "no", "be", "been", "heere", "shal", "quod", "whan",
"s", "were", "at", "wel", "oure", "have", "what", "or", "if", "hath", "on", "ne", "unto",
"us", "seye", "which", "men", "swich", "ful", "alle", "ther", "yet", "thou", "was", "upon",
"kan", "youre", "han", "thy", "hadde", "hym", "seyde", "thyng", "every", "hir", "they",
"er", "than", "out", "page", "good", "thee", "hem", "anon", "many", "greet", "an",
"sholde", "any", "moore", "right", "eek", "we", "tyme", "also", "doon", "lat", "therefore",
"noon", "do", "thus", "page", "incipit", "pars", "explicit", "t")
```

```
division_names = c("GP", "KnT", "MilP", "MilT", "RvP", "RvT", "CkP", "CkT", "MLI",
"MLP", "MLT", "MLE", "WBP", "WBT", "FrP", "FrT", "SumP", "SumT", "CIP", "CIT",
"MerP", "MerT", "MerE", "SqP", "SqT", "FranW", "FranP", "FranT", "PhyT", "PardI",
"PardP", "PardT", "ShT", "HW", "PrP", "PrT", "ThP", "Th", "HI", "Mel", "MkP",
"MkT", "KnI", "NPT", "NPE", "SNP", "SNT", "CYP", "CYT", "ManP", "ManT",
"ParsP", "ParsT")
```

```
tales = xml_find_all(read_html("Oxford_Canterbury_Tales.html"), "//div[@class='tale' or
@class='prologue' or @class='introduction' or @class='epilogue']"# or
@class='portrait']")
```

```
#Get all unique words used in the Tales
```

```
tales_vocab = xml_text(tales)
tales_vocab = unlist(tales_vocab)
temp_vocab = c()
for (v in 1:length(tales_vocab))
{
  tales_txt = unlist(strsplit(tales_vocab[v], split = "\\W"))
  temp_vocab = c(temp_vocab, tales_txt)
  temp_vocab = tolower(temp_vocab)
  temp_vocab = unique(temp_vocab)
}
```

```
tales_vocab = temp_vocab
tales_vocab = tales_vocab[!(tales_vocab %in% stop_words_chaucer)]
tales_vocab = tales_vocab[-grep("[0-9]", tales_vocab)]
tales_headers = xml_find_all(tales, ".//h3")
tales_txt = xml_text(tales)
```

```
#Build matrix. Each row is a word, each column is a text division. Each cell is the count
of that word in that division.
```

```
tales_matrix = matrix(0, nrow = length(tales_vocab), ncol = length(tales))
rownames(tales_matrix) = tales_vocab
colnames(tales_matrix) = division_names
for (i in 1:length(tales))
{
  print(i)
```

```

temp_txt = tales_txt[i]
temp_txt = unlist(strsplit(temp_txt, split = "\\w"))
temp_txt = tolower(temp_txt)
temp_txt = temp_txt[!(temp_txt %in% stop_words_chaucer)]
temp_txt = temp_txt[-grep("[0-9]", temp_txt)]
temp_txt = temp_txt[temp_txt != ""]
tales_freqs = table(temp_txt)
tales_matrix[names(tales_freqs),i] = tales_freqs
}

#Break up by sentences, create index for source tale for each sentence
all_sentences = c()
sentence_index = c()
for (i in 1:length(tales))
{
  print(i)
  temp_txt = tales_txt[i]
  temp_sentences = unlist(strsplit(tales_txt[i], split = "[.?!] |/"))
  temp_sentences = gsub("-- ", "", temp_sentences, fixed = TRUE)
  temp_sentences = gsub("\n", "", temp_sentences, fixed = TRUE)
  temp_sentences = gsub(" +", " ", temp_sentences)

  temp_sentences = tolower(temp_sentences)
  temp_sentences = trimws(temp_sentences)
  all_sentences = c(all_sentences, temp_sentences)
  sentence_index = c(sentence_index, rep(i, length(temp_sentences)))
}

# Convert tale numbers to tale labels
sentence_divs = division_names[sentence_index]

#### EOF ####

#### Canterbury Tales Analysis Phase 2 - Words - Sentences - Matrix ####

# Create word-sentence matrix
m = matrix(0, nrow(tales_matrix), length(all_sentences))
rownames(m) = rownames(tales_matrix)
colnames(m) = sentence_divs

for (i in 1:length(all_sentences))
{
  print(i)
  txt = all_sentences[i]
  txt = unlist(strsplit(txt, split = "\\W"))
  txt = txt[txt %in% rownames(m)]
}

```

```

txt = txt[txt != ""]
sentences_freqs = table(txt)
m[names(sentences_freqs),i] = sentences_freqs
}

# Convert to sparse matrix and build latent-semantic model
m_sparse = Matrix(m)
mp = ppmi(m_sparse)
ms = irlba(mp, nv = 100)
words = ms$u %*% diag(ms$d)
docs = ms$v %*% diag(ms$d)
rownames(words) = rownames(m_sparse)
rownames(docs) = 1:nrow(docs)

#### EOF ####

#### Canterbury Tales Analysis Phase 3 - Topic Lists ####

# Build word/topic list for whole text
kclusts = kmeans(words, centers = 50, iter.max = 30)
tales_groups = matrix("", 50, 100)
for (i in 1:nrow(kclusts$centers))
{
  print(i)
  vec = kclusts$centers[i,]
  sims = similarity(words, vec)
  tales_groups[,i] = paste(names(sims),round(sims,6))
}

# Build word/topic lists for individual divisions
knight_groups = semantic_model("KnT")
squire_groups = semantic_model("SqT")
wife_groups = semantic_model("WBT")

#Build sentence similarity lists for individual terms
keyword_terms_vector = c("arcite", "emelye", "knyght", "palamon", "theseus", "ypolita",
"chivalrie", "curteisie", "fredom", "honour", "trouthe", "worthy")

sentence_similarity_matrix = matrix(0,nrow=15,ncol=0)
temp_sim_matrix = matrix(0,nrow=15,ncol=2)
for(i in 1:length(keyword_terms_vector))
{
  temp_sim_matrix[,1] =
round(similarity(docs,words[keyword_terms_vector[i,]],6)[1:15]

```

```

temp_sim_matrix[,2] =
all_sentences[as.integer(names(similarity(docs,words[keyword_terms_vector[i],])))][1:15]
]
colnames(temp_sim_matrix) = c(keyword_terms_vector[i], keyword_terms_vector[i])
sentence_similarity_matrix = cbind(sentence_similarity_matrix,temp_sim_matrix)
}

#Build word similarity lists for those terms across the Tales as a whole
keyword_matrix = matrix(0, nrow = 15, ncol = length(keyword_terms_vector))
for(i in 1:length(keyword_terms_vector))
{
print(keyword_terms_vector[i])
sims = similarity(words,keyword_terms_vector[i],F,16)[2:16]
keyword_matrix[,i] = paste(names(sims),round(sims,6))
}

rownames(keyword_matrix) = c(1:15)
colnames(keyword_matrix) = keyword_terms_vector

#### EOF ####

```

APPENDIX B

SAMPLE OF RAW XML DATA

```
<div xml:id="knight" type="portrait"><head>The Knight's Portrait</head>
<l n="43">A knyght ther was, and that a worthy man, </l>
<l n="44">That fro the tyme that he first bigan </l>
<l n="45">To riden out, he loved chivalrie, </l>
<l n="46">Trouthe and honour, fredom and curteisie. </l>
<l n="47">Ful worthy was he in his lordes werre, </l>
<l n="48">And therto hadde he riden, no man ferre, </l>
<l n="49">As wel in cristendom as in hethenesse, </l>
<l n="50">And evere honoured for his worthynesse. </l>
<l n="51">At alisaundre he was whan it was wonne. </l>
<l n="52">Ful ofte tyme he hadde the bord bigonne </l>
<l n="53">Aboven alle nacions in pruce; </l>
<l n="54">In lettow hadde he reysed and in ruce, </l>
<l n="55">No cristen man so ofte of his degree. </l>
<l n="56">In gernade at the seege eek hadde he be </l>
<l n="57">Of algezir, and riden in belmarye. </l>
<l n="58">At lyeys was he and at satalye, </l>
<l n="59">Whan they were wonne; and in the grete see </l>
<l n="60">At many a noble armee hadde he be. </l>
<l n="61">At mortal batailles hadde he been fiftene, </l>
<l n="62">And foughten for oure feith at tramyssene </l>
<l n="63">In lystes thries, and ay slayn his foo. </l>
<l n="64">This ilke worthy knyght hadde been also </l>
<l n="65">Somtyme with the lord of palatye </l>
<l n="66">Agayn another hethen in turkye. <pb n="18"/></l>
<l n="67">And everemoore he hadde a sovereyn prys; </l>
<l n="68">And though that he were worthy, he was wys, </l>
<l n="69">And of his port as meeke as is a mayde. </l>
<l n="70">He nevere yet no vileynye ne sayde </l>
<l n="71">In al his lyf unto no maner wight. </l>
<l n="72">He was a verray, parfit gentil knyght. </l>
<l n="73">But, for to tellen yow of his array, </l>
<l n="74">His hors were goode, but he was nat gay. </l>
<l n="75">Of fustian he wered a gypon </l>
<l n="76">Al bismotered with his habergeon, </l>
<l n="77">For he was late ycome from his viage, </l>
```

<l n="78">And wente for to doon his pilgrymage. </l>
</div>

...

<div n="MilP" type="prologue"><head>The Miller's Prologue</head>

<l n="3109">Whan that the knyght had thus his tale ytoold, </l>

<l n="3110">In al the route nas ther yong ne oold </l>

<l n="3111">That he ne seyde it was a noble storie, </l>

<l n="3112">And worthy for to drawen to memorie; </l>

<l n="3113">And namely the gentils everichon. </l>

<l n="3114">Oure hooste lough and swoor, so moot I gon, </l>

<l n="3115">This gooth aright; unboked is the male. </l>

<l n="3116">Lat se now who shal telle another tale; </l>

<l n="3117">For trewely the game is wel bigonne. </l>

<l n="3118">Now telleth ye, sir monk, if that ye konne </l>

<l n="3119">Somwhat to quite with the knyghtes tale. </l>

<l n="3120">The millere, that for dronken was al pale, </l>

<l n="3121">So that unnethe upon his hors he sat, </l>

<l n="3122">He nolde avalen neither hood ne hat, </l>

<l n="3123">Ne abyde no man for his curtesie, </l>

<l n="3124">But in pilates voys he gan to crie, </l>

<l n="3125">And swoor, by armes, and by blood and bones, </l>

<l n="3126">I kan a noble tale for the nones, </l>

<l n="3127">With which I wol now quite the knyghtes tale. </l>

<l n="3128">Oure hooste saugh that he was dronke of ale, </l>

<l n="3129">And seyde, abyd, robyn, my leeve brother; </l>

<l n="3130">Som bettre man shal telle us first another. </l>

<l n="3131">Abyd, and lat us werken thriftily. </l>

<l n="3132">By goddes soule, quod he, that wol nat I; </l>

<l n="3133">For I wol speke, or elles go my wey. </l>

<l n="3134">Oure hoost answerde, tel on, a devel wey! </l>

<l n="3135">Thou art a fool; thy wit is overcome. </l>

<l n="3136">Now herkne, quod the millere, alle and some! </l>

<l n="3137">But first I make a protestacioun </l>

<l n="3138">That I am dronke, I knowe it by my soun; </l>

<l n="3139">And therefore if that I mysspeke or seye, <pb n="48"/></l>

<l n="3140">Wyte it the ale of southwerk, I you preye. </l>

<l n="3141">For I wol telle a legende and a lyf </l>

<l n="3142">Bothe of a carpenter and of his wyf, </l>

<l n="3143">How that a clerk hath set the wrightes cappe. </l>

<l n="3144">The reve answerde and seyde, stynt thy clappe! </l>

<l n="3145">Lat be thy lewed dronken harlotrye. </l>

<l n="3146">It is a synne and eek a greet folye </l>

<l n="3147">To apeyren any man, or hym defame, </l>

<l n="3148">And eek to bryngen wyves in swich fame. </l>

<l n="3149">Thou mayst ynogh of othere thynges seyn. </l>

<l n="3150">This dronke millere spak ful soone ageyn </l>

<l n="3151">And seyde, leve brother osewold, </l>
<l n="3152">Who hath no wyf, he is no cokewold. </l>
<l n="3153">But I sey nat therfore that thou art oon; </l>
<l n="3154">Ther been ful goode wyves many oon, </l>
<l n="3155">And evere a thousand goode ayeyns oon badde. </l>
<l n="3156">That knowestow wel thyself, but if thou madde. </l>
<l n="3157">Why artow angry with my tale now? </l>
<l n="3158">I have a wyf, pardee, as wel as thow; </l>
<l n="3159">Yet nolde I, for the oxen in my plogh, </l>
<l n="3160">Take upon me moore than ynogh, </l>
<l n="3161">As demen of myself that I were oon; </l>
<l n="3162">I wol bileve wel that I am noon. </l>
<l n="3163">An housbonde shal nat been inquisityf </l>
<l n="3164">Of goddes pryvetee, nor of his wyf. </l>
<l n="3165">So he may fynde goddes foyson there, </l>
<l n="3166">Of the remenant nedeth nat enquere. </l>
<l n="3167">What sholde I moore seyn, but this millere </l>
<l n="3168">He nolde his wordes for no man forbere, </l>
<l n="3169">But tolde his cherles tale in his manere. </l>
<l n="3170">M' athynketh that I shal reherce it heere. </l>
<l n="3171">And therefore every gentil wight I preye, </l>
<l n="3172">For goddes love, demeth nat that I seye </l>
<l n="3173">Of yvel entente, but for I moot reherce </l>
<l n="3174">Hir tales alle, be they bettre or werse, </l>
<l n="3175">Or elles falsen som of my mateere. </l>
<l n="3176">And therefore, whoso list it nat yheere, </l>
<l n="3177">Turne over the leef and chese another tale; </l>
<l n="3178">For he shal fynde ynowe, grete and smale, </l>
<l n="3179">Of storial thyng that toucheth gentillesse, </l>
<l n="3180">And eek moralitee and hoolynesse. </l>
<l n="3181">Blameth nat me if that ye chese amys. </l>
<l n="3182">The millere is a cherl, ye knowe wel this; </l>
<l n="3183">So was the reve eek and othere mo, </l>
<l n="3184">And harlotrie they tolden bothe two. </l>
<l n="3185">Avyseth yow, and put me out of blame; </l>
<l n="3186">And eek men shal nat maken ernest of game. </l>