“The Fire of Love and Joy of Chivalry”

A Lexical Frequency and Semantic Category Analysis of the Faerie Queene

Beth Beattie

August 2021

Submitted in fulfilment of the requirements for
the degree of Master of Philosophy

English Language and Linguistics

School of Critical Studies

College of Arts

University of Glasgow

ABSTRACT

While *The Faerie Queene* has been the subject of copious literary analysis, there has been little research done using quantitative corpus analysis techniques. This research aims to determine the core vocabulary and semantic categories of the first book of *The Faerie Queene* with the purpose of uncovering Spenser's key interests and motivations. There are two main stages to the project: lexical frequency and concordance analyses performed using WMatrix and AntConc, followed by a semantic category analysis using the Historical Thesaurus of English.

The results of these analyses found that the core vocabulary and semantic categories are mostly interlinked, focusing on words relating to chivalry and the human body. Examining the broader semantic categories, however, highlights the importance of emotion and social class, in addition to religion and morality. This indicates that despite the allegory of the book, Spenser adheres to the traditional themes of chivalric literature like love and social status. He also uses emotion and body part words to encode more abstract concepts within the text and uses suffering and morality to accentuate his interpretation of Protestantism found within the text. These findings link to what is known about Spenser, with particular focus on aspects of literary style and what it means to be a good Protestant in the sixteenth century.

ACKNOWLEDGEMENTS

First, I would like to thank my supervisors Jeremy Smith and Fraser Dallachy for their guidance over the course of this research. You have provided invaluable support throughout this challenging year and helped guide me towards becoming a better researcher.

I would also like to express my gratitude towards the English Language and Linguistics departments at the University of Lancaster and the University of Glasgow; the former for their work on programs and techniques used in this research, and the latter for their sense of community that helped provide encouragement.

This research would not have been possible without the support of my friends and family. I would especially like to thank Alex Jarvis for providing copious amounts of tea and putting up with me when I was most stressed.

Lastly, I would like to thank my cat Stevie Neeps, who did not help in any way at all.

CONTENTS

LIST OF FIGURES

1 INTRODUCTION

This research aims to gain new insights into Edmund Spenser's interests and cultural setting by applying modern corpus analysis techniques to *The Faerie Queene*. The core vocabulary of *The Faerie Queene* is determined by examining the most frequent lexical items, and their semantic categories are identified using the semantic tagger based on the Historical Thesaurus of English. This semantic distribution is the basis for establishing Spenser's main interests in writing *The Faerie Queene* as well as his setting within early modern culture.

This research benefits from advances in corpus linguistics by examining *The Faerie Queene* in a more systematic way than previous studies. Furthermore, it incorporates corpus analysis methods in a new way by combining more statistical analyses with the less concrete domain of semantics to work towards bridging linguistic and literary ideas.

Underpinning this thesis is a two-stage analysis of a corpus compiled from the first book of *The Faerie Queene*. The first stage consists of using corpus analysis software to determine the frequency of lexical items and the collocations of the most common lexemes. By contrast, the second stage focuses on semantic tagging and establishing the semantic categories Spenser employs the most often. These methods of analysis are available thanks to the development of more modern techniques in the field of corpus linguistics, which – by making use of these new approaches – allows us to supplement the wealth of previous literary and linguistic research on this text.

The Faerie Queene is arguably one of the most important works of Early Modern English literature and has thus received a lot of attention from scholars over the centuries. From a literary standpoint, examinations of the allegory and the mythological aspects of the poem have been conducted with some degree of depth (Tuve 1966; Craig 1972; Davis, W. 2002). The commentary on Elizabethan society within the poem has also been a subject of note (Green 1974). Scholarly interest does not stop with the poem and its thematic material; Spenser himself has been a frequent subject of historians' work, in particular his relationship with Ireland (Canny 1983; Maley 1994b) and his ties to other members of his discourse community.

For historical linguists, Spenser's language has been often discussed and debated- even Samuel Johnson had opinions on the archaisms within Spenser's works (Turnage and Spenser 1970).

The language that makes Spenser distinctive among poets has been the subject of debate, with a variety of conclusions being drawn about the extent and sources of archaic terms and borrowings found throughout Spenser's works (Parker 1925; Stephens 2010). While the origin of such words is not the focus of this research, it is important to bear in mind that Spenser may be evoking earlier senses to imbue his work with a unique literary flavour.

However, the overwhelming majority of research on Spenser is qualitative in nature, with a few exceptions. Osgood, in an impressive piece of pioneering analysis (1915), created an early concordance of *The Faerie Queene* with an immense level of detail, given the technological limitations of the time. However, with the resources now available, much more comprehensive analyses are now possible, providing more quantitative data on Spenser's *The Faerie Queene* and its linguistic contexts.

This research acts as a proof of concept for this form of corpus analysis, for which *The Faerie Queene* is the ideal candidate due to it having been written towards the end of Spenser's career once his literary style had matured. Focusing on one book allows for expansion to the entirety of *The Faerie Queene* in future work; the first book is the largest of the six that make up *The Faerie Queene* with 45,659 words, depending on the edition. It has also been noted that the first three books of *The Faerie Queene* have a larger vocabulary and contain more neologisms than the last three (Padelford 1941, pp.90–1), which makes any one of them all the more interesting to examine in depth.

In the corpus analysis, the focus is on establishing lexical frequencies before identifying the most significant semantic categories. The available corpus tools have the potential for many more types of analysis, but these are the most relevant to this research. Both work together in tandem, but examining the concordances of the most frequent lexical items is a necessary prerequisite for exploring the nuances of their semantic contexts; both analyses are needed to produce the desired outcome.

This research contributes to the existing body of work on *The Faerie Queene* by providing a different perspective on the lexical data. The use of quantitative methods of analysis is unlike the majority of previous methodologies that have been applied to Spenser and *The Faerie Queene* and facilitates a more thorough investigation of the choices in vocabulary that Spenser

made. By treating this research as a form of linguistic archaeology and methodically examining the available data, it becomes possible to corroborate or refute previously drawn conclusions with a greater body of evidence.

Furthermore, the incorporation of semantic categories alongside the other lexical analyses allows us to bridge literary and linguistic interpretations of *The Faerie Queene* in a new way. Relations to previous work across multiple domains, including literary analysis, can be made through finding statistical evidence for the strength of themes in the form of keyword frequencies. Such techniques are invaluable for future research pertaining to historical literature.

Modern methods of corpus analysis have already been proven to be beneficial in examining datasets from historical periods of English. What this research does is act as a pilot study for combining multiple types of analysis for potential future application to other larger sources.

The questions discussed in this research are:

1. Which words form the core vocabulary of *The Faerie Queene*?
2. Does this core vocabulary correspond with the core semantic categories?
3. What can this semantic distribution tell us about Edmund Spenser and his interests?

The methodology used to answer these questions is divided into two phases relating to the analysis of the corpus created using *The Faerie Queene* and a final phase interpreting the analysis with regards to Spenser himself. Before beginning analysis, however, the text of *The Faerie Queene* is run through spelling normalisation software (in this case, VARD 2) in order to standardise the spelling variation that is typical of Early Modern English and ensure the greatest degree of accuracy in subsequent stages.

Answering question one involves the use of WMatrix and AntConc to perform multiple analyses on the normalised corpus. Lexical frequency is the most significant method of determining the core vocabulary of the corpus, while examining the concordances of these words is beneficial for examining the contexts in which these words are used and establishing exactly how Spenser uses them.

To establish the most frequent semantic categories, the Historical Thesaurus Semantic Tagger is used to categorise each word in the first book of *The Faerie Queene* according to the hierarchy of the Historical Thesaurus of English. This categorisation structure is the best for using with historical texts and is the most comprehensive thesaurus of historical Englishes available. These categorisations of words in *The Faerie Queene* dataset will form the basis of analyses to determine the core semantic categories of the text. Finally, both the core vocabulary and core semantic categories, in addition to a closer examination of the text itself, will be used to determine what the first book of *The Faerie Queene* tells us about the ideas that Spenser found most interesting in writing the text.

The first book of *The Faerie Queene* follows the adventures of the Redcrosse Knight in his quest to win the hand of fair Una, encountering many foes typical to chivalric literature like dwarves, giants, and dragons. Literary analysis of the book has demonstrated the importance of themes such as religion (King, J.N. 1990; Mallette 1997), classical mythology (Craig 1972), and the supernatural (Wauchope 1903). Scholars have noted that while *The Faerie Queene* has many structural and thematic similarities with chivalric literature from previous centuries, the inclusion of references to contemporary issues like the Reformation and characters who represent real people known by Spenser makes *The Faerie Queene* an updated interpretation of the chivalric genre that better fits Spenser's needs and style.

Ultimately, this research finds that the core vocabulary of the first book of *The Faerie Queene* contains a lot of words relating to the genre of chivalric literature, such as "knight", "lady", "fair", and "foe". More interestingly, however, is the presence of body part words like "hand" and "eyes" among the most frequently used words in the text, in addition to the lack of words explicitly relating to religion, with the exception of "heaven". The core semantic categories feature many of the same concepts as the core vocabulary, but broadening the perspective to include higher tier HT categories highlights the importance of emotion, body parts, social class, religion, and morality. These categories reflect Spenser's interests in continuing the chivalric literary tradition, Protestantism, and encoding hidden meanings throughout the entire text. Spenser's inclusion of categories like love, death, and social class demonstrates his desire to maintain the use of traditional features of chivalric literature, despite the genre's manipulation to fit the allegory of the text. This allegory has been the subject of copious research (Tuve 1966;

Davis, W. 2002; Quilligan 2018), but in addition to the messages Spenser is communicating through the text, he is also encoding abstract concepts like ancestry and love through his use of words relating to body parts and emotions. Furthermore, suffering is used in tandem with concepts relating to morality to develop Spenser's depiction of Protestantism in this first book of *The Faerie Queene,* which reflects his views and interpretations of what it means to be a good Protestant in the late sixteenth century.

Chapter two is a discussion of the existing body of work on both *The Faerie Queene* and Spenser himself, in addition to exploring the existing research that uses corpus analysis techniques on works by Spenser and on literature more generally. Chapter three provides a description of the spelling normalisation process that was undertaken prior to the lexical analysis and semantic tagging discussed in chapters four and five respectively. Chapter six brings together all the analysis into a discussion of the most important ideas to Spenser before answering the posed research questions. Chapter seven concludes the thesis with a summary of the project and ideas for how this research could be expanded upon in future works.

2 RESEARCH CONTEXT

This chapter provides a summary of the body of work that already exists on Spenser, his discourse community, and *The Faerie Queene*. It begins with an overview of the key events in Spenser's life before *The Faerie Queene* itself is discussed, with the focus being on the language used within and the text's allegory. Finally, the existing concordances of Spenser's work will be explored, along with related work within corpus stylistics and studies from the Stanford Literary Lab that have a similar methodology to this research.

2.1 EDMUND SPENSER

Spenser was born in London in either 1552 or 1553 but his family likely came from the vicinity of Burnley. His immediate family is not certain, but he might be the son of an alderman or an ordinary journeyman (Hadfield 2021, p.1). Spenser attended Merchant Taylor's School, which was run by the humanist Richard Mulcaster, before matriculating at Pembroke College Cambridge in 1569. He is known to have been admitted as a sizar- "a poor scholar who earned his bed and board by performing a series of servant's duties" (Hadfield 2021, p.3), which cements the idea that Spenser's origins are at the less wealthy end of the still-developing middle class. Spenser ultimately graduated in 1576, having received both a bachelor's and master's degree, and began performing secretarial work for multiple powerful patrons like Robert Dudley, earl of Leicester, and Dr John Young, bishop of Rochester. His connections to such powerful people were most likely as a result of relationships fostered while at Cambridge, especially his friendship with the writer Gabriel Harvey (Hadfield 2014, p.89-92).

Spenser became the private secretary of Arthur, Lord Grey of Wilton in 1580, which precipitated his move to Ireland, where he would live for the majority of the remainder of his life. From his personal documents and financial records, it is clear that Spenser made full use of the cheap land and estates available to aspirational English settlers; he leased the manor of Enniscorthy, co.Wexford, in 1581, a ruined monastery in 1582, and the estate at Kilcolman in 1586, to name but a few (Hadfield 2021, pp.6–8). Spenser was directly affected by the effects of the Nine Years' War- in 1598, the Kilcolman estate was overrun by Irish rebels fighting against English rule in Ireland, forcing Spenser and his family to flee.

His work as a civil servant in multiple roles meant that Spenser travelled all over Ireland, as well as back to England on a number of occasions. All of Spenser's most notable works were published in England, mostly by Hugh Singleton, with the exception of *A View on the Present State of Ireland*, which – while completed in 1596 – was not published until after Spenser's death. *The Shepheardes Calendar* (1579) was published prior to his move to Ireland, *Complaints* (a collection of poems) was published in 1591, and *The Faerie Queene* in two editions in 1590 and 1596.

Spenser died in January 1599 in London after delivering letters to the privy council. He was buried in Westminster Abbey in Poet's Corner, with "his hearse being attended by poets, and mournful elegies and poems, with the pens that wrote them, thrown into the tomb" (Maley 1994a, p.80).

2.2 *The Faerie Queene*

The Faerie Queene is Spenser's largest and arguably most well-known work. Of the twelve planned books, six completed books and a handful of cantos of a seventh were published- the first three in 1590, with revised versions being included with the latter three in 1596.[1] In the 1596 edition, *The Faerie Queene* is appended by a variety of sonnets dedicated to a number of important figures of the period, as well as prefaced by a letter written by Spenser to Sir Walter Raleigh explaining the role of allegory within the "poet historical" (Hadfield 2021, pp.10–11). *The Faerie Queene* is often described as a romantic epic with heavy influence from chivalric literature of the Matter of Britain. Such origins of *The Faerie Queene*, as well as other key aspects of the romantic epic applicable to the poem, are discussed in Wauchope's introductory notes to the 1903 edition of *The Faerie Queene*:

> "The scene of the adventures is laid in the enchanted forests and castles of the far away and unreal fairyland of mediæval chivalry, and the incidents themselves are either highly improbable or frankly impossible. The language is frequently archaic and designedly unfamiliar. Much of the machinery and properties used in carrying on the

[1] The 1596 edition is the edition used in this research, available through *Early English Books Online* at http://name.umdl.umich.edu/A12778.0001.001.

> story, such as speaking myrtles, magic mirrors, swords, rings, impenetrable armour, and healing fountains, is supernatural. All the characters—the knights, ladies, dwarfs, magicians, dragons, nymphs, satyrs, and giants—are the conventional figures of pastoral romance" (1903, p.8).

The weaving together of chivalry, the pastoral, and the supernatural forms the basis for all six books of *The Faerie Queene*, with the cast of main characters deriving from Arthurian legends. While characters and events link together across books, book one is considered a mostly self-contained story, which – combined with its length – make it the most appropriate book for this research (Greenblatt 2018, p.715). The first book tells the story of the Redcrosse Knight – Saint George – as he embarks on multiple quests culminating in saving his love Una's parents by defeating a dragon. Along the way, he encounters many foes typical of chivalric and pastoral literature like giants, dwarves, and corrupted knights, in addition to being repeatedly seduced by Una's evil counterpart Duessa. Ultimately, the Redcrosse Knight rescues Una and swears to marry her, but is recalled to the service of Gloriana – the representation of Queen Elizabeth – before they say their vows.

The Faerie Queene has been the focus of literary and linguistic analysis for centuries, ever since its initial publication. There are a plethora of aspects that could be studied: the role of women (Anderson 2018); ties to classical mythology (Craig 1972); social commentaries (Green 1974); and where the poem fits within Protestantism and the Reformation (King, J.N. 1990; Mallette 1997). For the purposes of this project, I shall focus on features of Spenser's language and the allegory within *The Faerie Queene*.

It has been identified that Spenser's literary style has four main features: archaisms, dialect words, borrowings, and neologisms (Wauchope 1903, p.12; Stephens 2010, pp.371–4). Of these four, he is most well-known for his use of neologisms and archaic language. Archaisms are typical of pastoral literature, and a strong influence on Spenser's archaic language is Chaucer (Parker 1925, p.81; Higgins 1990, p.17; King, A. 2010, p.554). A context in which a large number of Chaucerisms or more general archaisms is clear is in one of Spenser's earlier works *The Shepheardes Calender*, where "leasure" (opportunity) and "pleasaunce" (joy, pleasure) are used, in addition to others (Parker 1925, p.84). This tradition continues into the first book of *The Faerie Queene*, where "keepe" (heed, give attention to) and "effraide"

(frightened) are among the archaisms used. There are even direct references to Chaucer in the text of the fourth book of *The Faerie Queene*, so much was his influence on Spenser (Higgins 1990, pp.18–19).

The amount of archaic language in Spenser's style is not uniquely interesting to modern scholars- it was a common topic of discussion for those who read Spenser's work at the time and shortly following. The mysterious commentator on Spenser's works "E.K." comments on the use of archaisms in *The Shepheardes Calendar* and their usewas a common criticism among reviewers of *The Faerie Queene* upon publication (Stephens 2010, pp.371–2). However, there were those who viewed Spencer's archaic language in a more positive or useful light. When searching for quotations for his *Dictionary of the English Language*, Samuel Johnson used Spenser's work as source material for 2,878 entries, of which 1,520 are drawn from *The Faerie Queene* (Turnage and Spenser 1970, p.559). Regardless of scholarly views on the usefulness of such words, Spenser seems to have used archaisms to emphasise references to the pastoral with "rustic" language (McElderry 1932, p.149). It is therefore clear that archaisms are an important and defining aspect of Spenser's literary style as a means of enhancing key themes in his work.

There has been a lot of scholarly interest in Spenser's neologisms and which words are first attested in his works (McElderry 1932; Padelford 1941), with some debate as to whether these first attestations are accurate and representative of Spenser's style (Gans 1979, p.379). However, it is still beneficial to examine how Spenser creates and changes words to suit his needs. There are five ways in which Spenser invents new words: altering the meanings of existing words, changing a word's part of speech, shortening or lengthening words, compounding, and creating pseudo-archaisms (Stephens 2010, p.373). "Dernly" was typically used to mean "secretly", but Spenser changed the meaning to "dismally" in his usage of the word. He also used "comprovincial" as an adjective, meaning "of or belonging to the same province", instead of a noun, which was previously standard. Words like "advance" would be shortened to "vauce" and affixes added to other words to create "recomfortless" and "picturals". Spenser also harked back to earlier forms of English by creating compounds like "undersong" and "love-lavish", as well as creating new words that sounded archaic, of which "ygoe" and even "faerie" are good examples (Stephens 2010, pp.373–4). It is worth noting,

though, that Gans argues that neologising could be viewed as a form of archaism in Elizabethan times as a way of imitating the styles of writers like Chaucer and Lydgate (1979, p.379). Given Spenser's interest in Chaucer, it is reasonable to treat archaisms and neologisms as related and not completely separate them into different features within Spenser's literary style.

Now that key aspects of Spenser's style have been established, the role of allegory in *The Faerie Queene* can be explored. In the sixteenth century, Protestants were of the opinion that literature should provide moral guidance in addition to being a source of entertainment, and *The Faerie Queene* is no exception to this belief (Wauchope 1903, p.9). Spenser achieves this through allegory, which can be described as a "literary technique of suggesting more than one meaning or sense, a matter of *extended* meaning rather than 'hidden meaning'" (Davis, B.E.C. 1933, p.152). Tuve furthers this by saying that "we allow [*The Faerie Queene's*] major fictions and figures to speak with a double voice" (1966, p.49), which means that everything and everyone in the poem can be treated as a symbol or representation of something else.

Most of the focus of research on the allegory of *The Faerie Queene* focuses on aspects of morality and religion. Wauchope divides the allegory into moral, religious, and political classes before listing what each character represents in each of these categories (1903, pp.9–11). In relation to the first book, Davis details the religious symbolism that runs deeply throughout the book and compares it to that of Dante's *Divine Comedy* and Langland's *Piers Plowman* (2002, pp.155–6). Hadfield is more specific with his interpretation that the allegory reflects the societal and religious difficulties following the Reformation and how the past can be accommodated in the new world being created (2011, pp.35–6). This observation brings language and allegory together, as by using archaic language and neologisms, Spenser is building on this idea of merging old and new both at the linguistic and thematic levels.

2.3 Similar Studies

Of the research that makes use of corpus analysis methodologies with regards to Spenser, the major works focus on creating concordances of *The Faerie Queene* and his other works. The definitive work of this type is Osgood's *Concordance to the Poems of Edmund Spenser* (1915), which features an almost-one-thousand page concordance of every word used by Spenser in every work he ever published. For *The Faerie Queene* citations, Osgood chose to use a

combination of two versions of the text published at the beginning of the twentieth century, both based on the 1596 edition, and included each item of varied vocabulary in the list where possible (1915, p.vii). With regards to spelling variation, he included each spelling variant separately, but links each variant to a single headword with the concordances listed there- an effective way of acknowledging spelling variation but without unnecessarily complicating the data. The purpose of this work is to be a data source which researchers could refer to for further research; it contains no analysis on the data it records. Having such a repository would have been invaluable prior to computer programs that gather data in this manner at the drop of a hat, but despite its seeming outdatedness, the *Concordance* is still worth considering in modern research on Spenser and related subjects (Quilligan 2018; Walls 2020).

Osgood's work is not the only concordance of *The Faerie Queene*; Yamashita et al.'s Comprehensive Concordance to the Faerie Qveene 1590 (1990) instead focuses on using the first three books to create a raw data source for studying *The Faerie Queene*. The decision to concentrate on books one to three stems from the fact they use the 1590 edition as source material, which only includes books one to three of *The Faerie Queene*. Unlike with Osgood's work, each spelling variant is treated as a headword, facilitating analyses that focus on spelling variation, but homographs are merged into a single entry (so "left" encompasses both nouns and adjectives). Yamashita et al. take things a step further than Osgood by including frequencies for every word listed. Technological advancements make such developments possible – the concordance is computer-generated – and are also demonstrated by the inclusion of CDs with facsimiles of *The Faerie Queene* source text and resulting data alongside the book. The purpose of the book is similar to that of Osgood's- a data source. The Comprehensive Concordance does not contain any analysis of the compiled data, and even though a companion book was published a few years later, it contains additional data like comparisons between specific books and notes on spelling variation and differences between editions (Yamashita 1993). There is not much in the way of further analysis of this data in other research, though, except in reference to a discussion of allegory within the broader early modern period (Sandberg 2016) and Brown and Lethbridge's work discussed below (2021). This is probably due to the development of software not long after the publication date that

allowed for researchers to create their own datasets easily, with little need to use pre-existing datasets as the sole source materials for a project.

There are still uses for creating concordance lists, however, as long as there is a specific purpose to them. Brown and Lethbridge's *Concordance to the Rhymes of The Faerie Queene* (2021) builds on Brown's earlier work on rhyme and poetic structure within *The Faerie Queene* (2019) by creating a list of every word in the text alongside its rhymes. The list is supplemented by a section listing the data in various ways, including alphabetically and by frequency, and another detailing noteworthy features about Spenser's rhymes, including specific syntactic structures and epithets. Brown and Lethbridge used every edition of *The Faerie Queene* published during Spenser's lifetime, in addition to the *Mutabilitie Cantos* from the 1609 edition, stating that this provides the broadest and most accurate concordance of words attributed to Spenser. However, not every spelling variant is included in the actual list of rhymes, in order to prevent "dilution" (2021, p.xiii). This newest edition of the book is too new for it to have received many citations, but since rhymes depend more on pronunciation than spelling, identifying them is harder to do automatically via computer software, so focusing on rhymes makes this concordance more relevant to modern research than just a simple word or concordance list.

From a methodological perspective, this project's aims align with the field of corpus stylistics, which Mahlberg argues "can make use of innovative descriptive tools that not only fit into linguistic frameworks but also leave room to account for individual qualities of texts and thereby link in with literary interpretation" (2007b, p.219). This combination of literary and linguistic techniques is demonstrated in Mahlberg's work on Dickens, which also identifies word clusters as an important linguistic unit in frequency and keyness analyses (Mahlberg 2007a, 2010). Similar work has been performed on the language of Shakespeare, with Culpeper utilising a combination of keyword, part-of-speech, and semantic domain analyses – available in WMatrix – to examine the language of characters in *Romeo and Juliet* (2009). The use of WMatrix to perform these analyses is also demonstrated in Mahlberg and McIntyre's work on *Casino Royale*, with the additional element of grouping keywords by "fictional world signals" and "thematic signals" according to how the keywords function in the text (2011, p.209). While there is a significant body of work within corpus stylistics that uses lexical and semantic

analyses on works of literature, the semantic element uses the USAS tagset, often through the use of WMatrix (see chapter 4). Using the Historical Thesaurus Semantic Tagger to perform the semantic analysis in this project offers an alternative view on the semantic structure of *The Faerie Queene* that is missing from the current work in corpus stylistics.

Further research with similar aims and objectives to this project is also being completed at the Stanford Literary Lab. This research collective states its purpose as "[applying] computational criticism, in all its forms, to the study of literature" (Algee-Hewitt 2021), which also describes the methodology of this research. The Literary Lab have produced a number of publications, but the work that has the most ties to the subject matter of this thesis is the pamphlet series the lab produces. There have been seventeen of these short papers published by the lab,[2] each on a different topic in the vein of the Literary Lab's interests, of which four are relevant to this project.

The Emotions of London (Heuser, Moretti and Steiner 2016) builds on existing work on genre using a corpus of five thousand novels published between 1700 and 1900 that are set in London. The main portion of this project uses a type of "sentiment analysis" to create maps of locations within London that are associated with particular emotions. This sentiment analysis uses a program trained with a dictionary of words that are assigned a positive or negative value, comparing words found in a given text with this dictionary to assign words, passages, and therefore locations as "frightening", "happy", or somewhere in between. There were some discrepancies between the program and human interpretations: the crowd-sourced human taggers marked 12% of the passages as frightening, compared with the 1% of the program (Heuser, Moretti and Steiner 2016, p.6). Furthermore, the researchers acknowledge that the program's training using data from the Wall Street Journal makes it not ideally suited to historical texts. These issues surrounding program training and suitability to historical texts are relevant to this research, in addition to the use of semantics in discussing how the bounds of fictional London in literature did not expand with time as with fictional Paris (Heuser, Moretti and Steiner 2016, pp.4–5). However, the data used for the study is much later than the early

[2] As of August 2021.

modern period of *The Faerie Queene*, plus the study operates on a larger scale while focusing on significantly fewer aspects of the text than this thesis.

Another example from the Literary Lab of analysing large-scale text corpora is Moretti and Pestre's research on the language of the yearly reports of the World Bank (2015). Using reports dating from 1946 and 2012, the study examined changes in vocabulary, grammar, and semantic categories across the dataset, as well as the potential societal reasons behind these changes. Lexical frequency analysis was used with keywords within domains like financial language and management discourse, as well as with grammatical features like nominalisation and noun/verb ratios. This aspect of the methodology is comparable with this research on *The Faerie Queene*, despite the source data for the Literary Lab project not deriving from works of fiction. Of all the Literary Lab pamphlets, this project seems to be the most closely related to corpus linguistics, given the focus on the lexical and grammatical aspects of the World Bank corpus. However, there are no comparisons made with other similar material, and frequency of features is the only type of analysis performed, giving the research a rather limited viewpoint.

A project that also has a focus on the grammatical features of a corpus is Allison et al.'s *Style at the Scale of a Sentence* (2013). They use the Chadwyck-Healey corpus of nineteenth century texts as a basis for analysing clause types and functions. Furthermore, lexical frequency is incorporated into an examination of the most frequent words in a particular type of clause, and the clause types are mapped onto the semantic space of the corpus. They even identify specific verbs that are found more in particular genres like Gothic and Jacobin novels. While there are aspects of the methodology of the project that are similar to the methodology applied to *The Faerie Queene*, the focus on clauses and closed-class words is too grammatical to be replicated using the tools available for this project.

The final Literary Lab pamphlet that will be discussed here is Heuser and Le-Khac's Quantitative Literary History of 2,958 Nineteenth-Century British Novels (2012). As is indicated in the title, the project traced the concretisation of language in novels throughout the nineteenth century and the effect this has on narrative characterisation, and setting. The project also developed quantitative computational methods for tracking these changes, the most interesting of which is the "correlator". This tool takes an input "seed word" and outputs the words from the corpus

that follow the same trends in usage across time. Such a tool would be useful if this research was a diachronic examination of Spenser's works over the course of his life, but it is not designed for synchronic examinations. In order to ensure the relevancy of the outputted data to the seed words in Heuser and Le-Khac's study, the Historical Thesaurus taxonomy was used. They found the categorisation system perfectly suited to identifying relevant semantic categories within their data, as is the case with this research on *The Faerie Queene*, as well as expanding on the correlator's output (Heuser and Le-Khac 2012, p.7).

There are methodological similarities between this thesis and the Literary Lab projects. Using programs to tag texts according to emotional positivity or negativity, as well as finding words that have similar patterns of usage, relate to the semantic tagging and analysis portions of this research. Furthermore, the incorporation of lexical frequency acts as an important foundation on which additional analysis can be performed. However, the main differences between the Literary Lab research and this project are the time period and scope. The earliest texts used in the Literary Lab project date from the 1700s, whereas *The Faerie Queene* is over one hundred years older. Also, the extended time periods of these projects mean that they are diachronic in nature, as opposed to the synchronicity of this research. Lastly, each of these pamphlets focuses on one small area of analysis, as is to be expected from a smaller more informal format. While they cannot be expected to cover the same breadth of analysis of their data as a thesis, there is very little development of their ideas in other works beyond occasional references. Each project has so much potential for further exploration that is missing from the Literary Lab's body of work.

3 PREPARING THE TEXT

This chapter describes the spelling normalisation process that was undertaken prior to the data gathering stages. First, there is a description of how VARD2 – the spelling normalisation tool selected for this research – works, followed by the observations made during the training and tagging process.

3.1 ABOUT VARD

The variation in spelling during the early modern period makes corpus analysis challenging; each variant spelling is categorised as a different lexical item by current software. In order to rectify this, the spelling conventions in *The Faerie Queene* were normalised before any further analysis took place. The tool used to complete this was VARD2 (VARiant Detector), the second iteration of VARD, managed by Alistair Baron (Rayson, Archer and Smith, N. 2005; Baron and Rayson 2008; Archer et al. 2015).

There are three main features of VARD2: first, a text can be uploaded and spelling can be normalised individually by the user in the interactive mode; second, a batch mode, in which multiple texts can be normalised automatically; and lastly, VARD 2 can be trained by the user using alternative data, resulting in applications beyond the early-modern English of the base

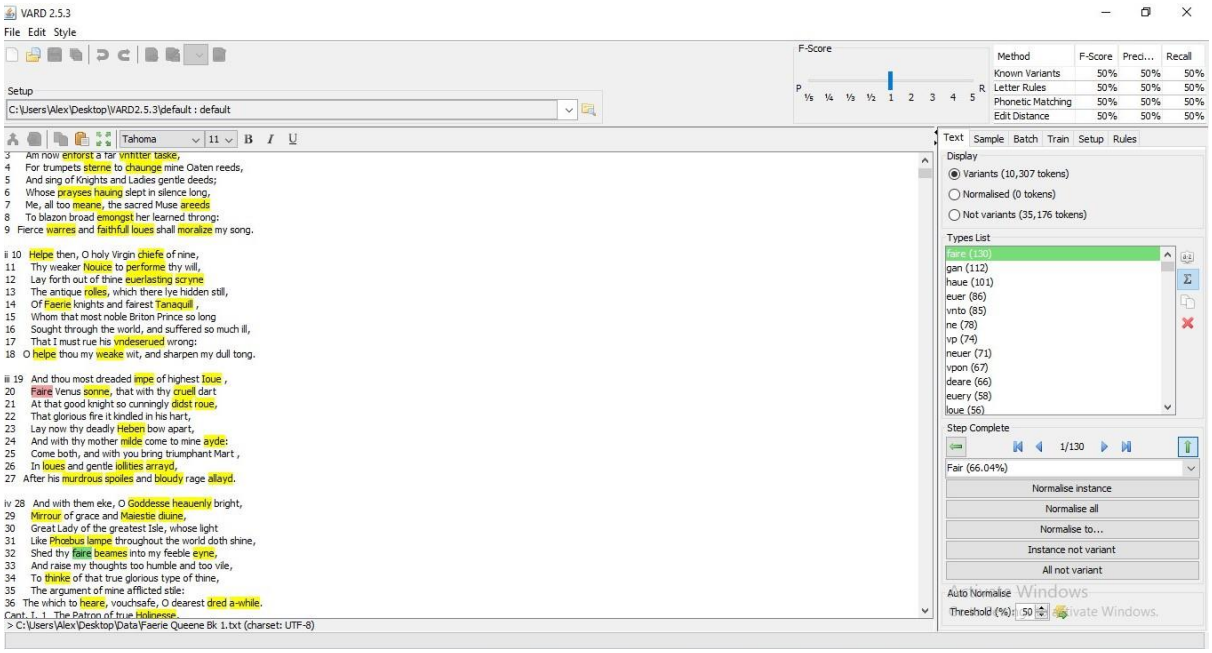


Figure 1: Interactive mode of VARD 2

programming. For the purposes of this research, only the interactive mode was relevant, which is the section of VARD 2 focused on here.

The screenshot shown in Figure 1 shows the interactive user interface of VARD 2. This segment of the program allows the user to manually normalise an uploaded text. VARD 2 automatically tokenises the text using the existing early-modern English training data (unless otherwise specified by the user) and marks each lexical item as "variants" or "not variants"; spellings marked as "variants" are highlighted within the text and listed in the "Text" pane. Upon selecting a variant spelling, the most likely standard spellings according to the training are listed, ranked by overall confidence. This allows the user to choose to normalise that individual occurrence of the spelling or all occurrences of this variant throughout the text. If the preferred spelling is not listed, it can be manually entered, and if the variant spelling is in fact not a variant, it can be manually marked as "not variant". Another option available to the user is to merge words commonly separated in Early Modern English (such as *to morrow*). Each selection made by the user affects the accuracy of the training, with the overall confidence, precision (the percentage of normalisations which are correct), and recall (the percentage shown in the top-right of the screen). A larger version of this table is shown in Table 1).

Method	F-Score	Prediction	Recall
Known Variants	50%	50%	50%
Letter Rules	50%	50%	50%
Phonetic Matching	50%	50%	50%
Edit Distance	50%	50%	50%

Table 1: Default confidence values in VARD2

In order to calculate the confidence scores upon which replacement spelling suggestions are given, the prediction and recall scores are combined into "f-score". By default, the weighting between prediction and recall is 50/50, but this may be changed by the user; in the case of this research, neither figure was given priority. On the y-axis, four methods for identifying variants are listed: "known variants" relates to known variant spellings for specific lexical items in VARD's training files, with prediction and recall being affected by how often these known variants are correctly identified; "letter replacement" uses the rules list file to identify known rules for letter replacements to create the intended word; "phonetic matching" uses an

adapted version of the Soundex phonetic matching algorithm to find the desired replacement spelling; and "edit distance" refers to the normalised Levenshtein distance used to rank each spelling suggestion offered by VARD 2 (Baron 2013).

Once normalised, the text can be exported with or without XML tags. These tags detail the changes made to each word, such as:

Any normalisation, either made automatically or by the user:

- <normalised orig="[variant form]" auto="[true/false]">
  [normalised form]</normalised>

Any word manually marked as "variant" by the user:

- <variant>[word]</variant>

Any word manually marked as "not variant" by the user:

- <notvariant>[word]</notvariant>

Multiple words merged into a single word:

- <join orig="[old string]">[new string]</join>

These tags can be used in creating alternative training for VARD 2, in addition to providing additional data for other corpus analysis software. For the purposes of this research, however, the tags were not necessary for subsequent analysis and were thus not included in the version of *The Faerie Queene* used in later analyses.

3.2 Observations

VARD's initial assessment of the first book of *The Faerie Queene* indicated that there were 10,307 tokens marked as variant spellings, compared with 35,176 non-variants. After manual correction, the number of variants that required normalisation came to 9,781, with 526 of the spellings originally marked by VARD as variants being reclassified as non-variants. Some of these words included names of characters, such as Duessa and Sansfoy; others were words that had just had the misfortune to not be included in VARD's original dictionary. This will be discussed in greater detail later.

Method	F-Score	Prediction	Recall
Known Variants	92.05%	98.73%	86.22%
Letter Rules	57.94%	46.68%	76.36%
Phonetic Matching	4.17%	2.13%	89.92%
Edit Distance	5.8%	3%	84.36%

Table 2: VARD confidence values post-normalisation

Throughout the manual normalisation process, the confidence scores given by VARD were automatically updated based on user input; Table 2 illustrates the scores given following the completion of the spelling normalisation of the first book of *The Faerie Queene*. While the recall values are all upwards of 75%, indicating a reasonable ability by VARD to adapt its training based on input, looking at each method more closely gives us an insight into the spelling systems in Spenser's work. First, the extremely high prediction score with regards to known variants shows that the vast majority of variant spellings used by Spenser were found on VARD's variant spellings list; there were very few forms wild enough to be excluded from that list. However, looking at the letter rules, the f-score was only around 58%. This suggests that the letter mappings between Spenser's and present-day English are a lot more diverse than those encoded into VARD. Phonetic matching and edit distance are much less relevant with regards to forming any conclusions about Spenser's spelling, but variation from the early-modern norm can already be determined.

As the spelling normalisation process progressed, there was a marked increase in the number of manual corrections to the spellings VARD suggested. The accuracy of suggested forms decreased dramatically alongside word frequency - so much so that the correct spelling had to be typed in manually for almost every unique instance of a spelling form. This is to be expected given the variability of Early Modern English spelling and the limitations of VARD's dictionary and training. However, Spenser's fondness for archaic vocabulary resulted in a number of words being used in multiple instances that were not found in VARD's dictionary and are not typically familiar to modern readers. This resulted in using the Oxford English Dictionary (OED) to identify obsolete words without recognisable modern-day reflexes, and then using the OED headword form for these items where they could be identified. This method helped to preserve words like *dight* "to equip, fit out, furnish" and *aread* "to counsel, advise" that might have

otherwise been misinterpreted (dight, v. 2020; aread | arede | areed, v. 2020). While it is possible to add words to VARD's dictionary in manual normalisation mode, the decision was made not to do this because the words were too rare to make such additions worthwhile.

In additional to curious lexical items, examining *The Faerie Queene* in VARD illuminated potentially noteworthy features of the style of the text that would have gone unnoticed in an already normalised version of the text. Regardless, the striking spelling characteristics in this edition of *The Faerie Queene* are worth discussing. It is of course unknown whether these features are Spenser's own or belonging to the publisher or printer; there are no surviving copies of Spenser's literary works written in his own hand with which to compare the printed book (Beal 2013). Whether deliberate or unintentional, there is a lack of internal consistency in the spelling system used in this text: "pity" fluctuates between <pittie> and <pitty>, "also" is sometimes realised as <als> or <alsoe>, and "ward" is sometimes spelled with a <g> in the manner of Middle French. There are also variations in how Spenser treats grammatical aspects. The text varies between plurals ending in <-s> and <-n>; "foes" is also found as "fone", and "eyes" fluctuates between "eies", "eyne", and "eyen".[3] There are also several instances where the northern suffix <-and> is used instead of the more typically southern <-ing>. This feature in particular might reflect Spenser's use of archaic language to contribute to the rustic and pastoral atmosphere of the text. If these spellings can be attributed to Spenser himself, it is possible that these variations in spellings not only relate directly to his experimental views towards language, but also Spenser's own idiolect, given his exposure to both northern and southern dialects. By taking the time to normalise such spelling variation, not only are further analyses easier and more accurately performed, but an insight into Spenser's language is also gained from the outset.

[3] For the purposes of subsequent analyses, the variations in plural forms were normalised to the present-day form, eg. "eyes" and "foes".

4 CORE VOCABULARY: LEXICAL ANALYSIS

This chapter details the process of performing the lexical analysis on *The Faerie Queene* dataset. After introducing the software used for these analyses, the methodology behind gathering the lexical frequency with WMatrix is discussed, followed by a brief discussion of the data and initial observations made. This structure is then repeated for the collocation analysis performed in AntConc, with the chapter concluding with a discussion of the limitations in using this methodology.

4.1 ABOUT WMATRIX AND ANTCONC

The dataset deriving from the first book of *The Faerie Queene* was developed using two pieces of corpus analysis software: WMatrix4 and AntConc. Both programs have the ability to perform analyses of keywords, lexical frequencies, collocates, concordances, and n-grams, but each program has slightly different ways of executing these functions, so using a combination of the two programs provided the clearest outlook on the corpus data.

WMatrix4 is the most recent version of the WMatrix corpus analysis software released in 2018. Initially developed as a part of the REVERE project (Rayson et al. 2000), WMatrix incorporates the USAS semantic tagger (Rayson et al. 2004) and the CLAWS part-of-speech tagger (Leech, Garside and Bryant 1994) with more standard corpus analysis features like lexical frequency and collocation into a single web interface. As it is an online tool, each text must be uploaded to WMatrix and can be run through the tag wizard, which automatically tags parts of speech using the CLAWS7 tagset and semantic categories using the USAS tagset. This can be overridden if there is additional dictionary data which can be added, but this was not relevant for this research. Once uploaded and tagged, there are a wide variety of tools available for different analyses thanks to the incorporated semantic and part-of-speech taggers (see figure 2).

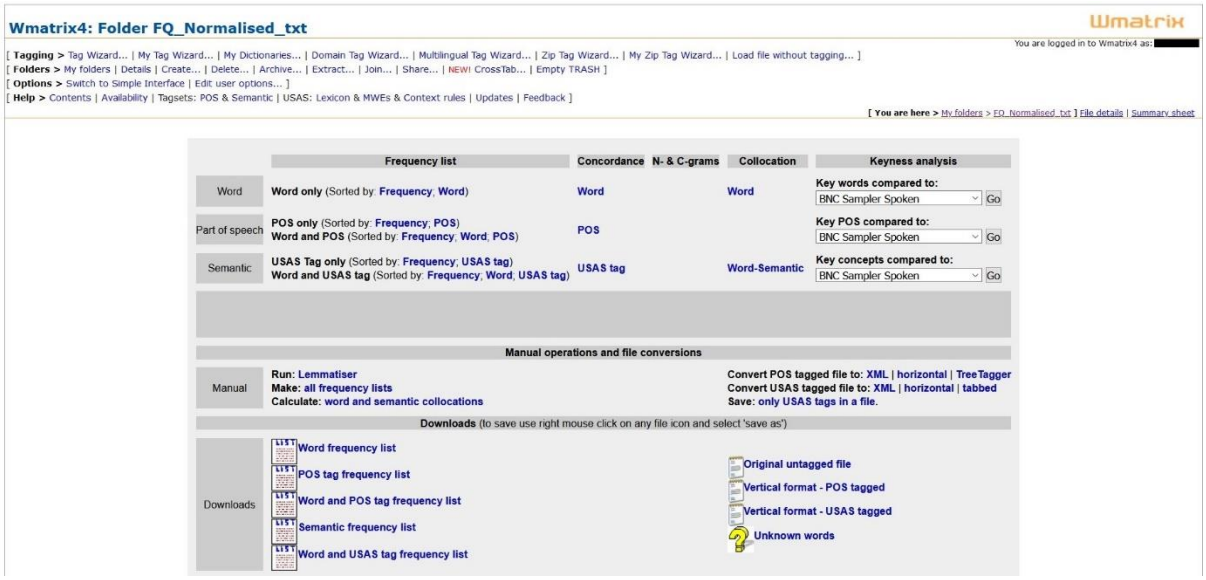


Figure 2: Main page of WMatrix4 after uploading a text

The main types of analysis that can be performed in WMatrix are lexical frequency, concordances, n-grams and c-grams (collapsed grams), collocations, and keyness compared with a reference corpus. Each of these analysis types can be performed on words, parts of speech, and semantic tags. All of these tools work in the browser, with data having to be copied and pasted elsewhere manually, but it is possible to download frequency lists as text files.

AntConc is another free corpus analysis tool used to develop *The Faerie Queene* corpus. The version used in this thesis is 3.5.9, released 2020, provides a reliable way to perform data analysis offline (Anthony 2019). In order to use AntConc, the relevant text must be opened and run through the 'Word List' tool before any further analysis can be performed. From there, it is possible to use the following tools:

- Concordance Tool
- Concordance Plot Tool
- File View Tool
- Clusters/N-Grams
- Collocates
- Keyword List

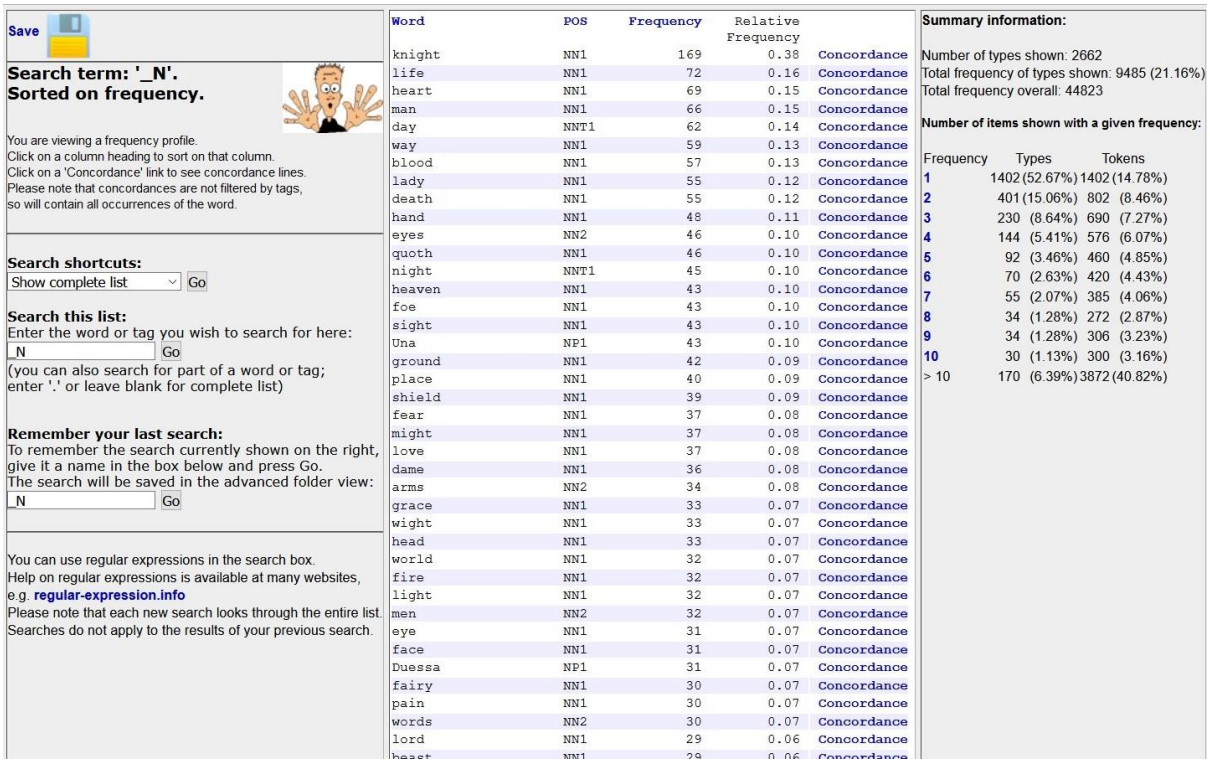


Figure 3: Noun frequency list in WMatrix

The Concordance tool performs KWIC (KeyWord In Context) searches in order to show the contexts in which words or phrases appear in the text. If you want to see the distribution of a word or phrase across the text, the concordance plot tool uses a 'barcode' format to illustrate the spread of usages within the text. As it suggests in the name, the file view tool allows the text to be viewed within AntConc. Clusters and N-Grams is the best tool to use to identify commonly-used phrases of a specific length and is similar in function to the concordance tool. Collocates identifies the most frequent words within a given frame of reference to a specific word, and finally, the keyword list is useful for comparing frequencies with a reference corpus.

4.2 WMatrix

4.2.1 Data

The most relevant analyses for this research available in WMatrix were the word and part-of-speech frequencies, word concordances and collocation, and word keyness compared with a reference corpus; the concordancing and collocation software within WMatrix, however, was still in beta at the time of research, so these analyses were performed with AntConc to ensure accuracy. Each of the frequency lists were easily obtained with WMatrix, facilitated by the

ability to filter words by part of speech and thus exclude words such as 'and' and 'the' (see figure 3). The parts of speech selected were nouns, verbs, and adjectives- all open-class categories with the potential to contain items belonging to a variety of semantic categories.

Each of the datasets corresponding to the main parts of speech chosen were recorded in an Excel spreadsheet, detailing the lexeme, CLAWS tag, raw frequency, and relative frequency. The relative frequency is the percentage frequency of the lexeme, which is calculated by dividing the lexeme's raw frequency by the total number of words in the corpus, then multiplied by 100.

The keyword analysis was more challenging to perform as it required a reference corpus. WMatrix already has access to various subsets of the British National Corpus, in addition to the British English 2006 and American English 2006 corpora. However, since the sources for these corpora date from the late twentieth century, they are not a useful reference point for comparing the Early-Modern English of *The Faerie Queene*, thus a new reference corpus had to be created. Early English Books Online (EEBO) provided the source material for the twenty-five texts used in this reference corpus, with each text dating between 1555 and 1647– around fifty years either side of the text's approximate date of composition – and being somewhat thematically similar to *The Faerie Queene*. Each text input into VARD to normalise the spelling and the resulting corpus is approximately four times the length of book one of *The Faerie Queene* (at 188,000 words compared with the first book of *The Faerie Queene's* 46,000), which provided a more appropriate reference point for the keyword analysis.[4] While the ideal size of a reference corpus is five times the size of the original corpus (Berber-Sardinha 2000, p.12), there was limited time available for normalising the spelling and part-of-speech tagging the texts within the reference corpus. Further research building on this pilot study would include a larger reference corpus for an even more accurate keyness analysis.

WMatrix uses the Log-Likelihood statistic (LL) to measure keyness (Rayson and Garside 2000), with an LL value over 6.63 being significant at $p<0.01$. The Log Ratio value is also given to help visualise the difference in relative frequency between the two corpora (Hardie 2014). These

[4] For more information about the texts included in the reference corpus, see Appendix A.

two values were the reference points used to determine the keywords of the first book of *The Faerie Queene*, the results of which were recorded in a spreadsheet.

Using this methodology to gather frequency data in WMatrix was one of the simplest ways to obtain the desired information. The user interface is clear and the tool allows for the search customisation required for this project, such as focusing on specific parts of speech. Furthermore, the level of detail given by including the relative frequencies of each word and the exact CLAWS part-of-speech tag adds the potential for additional perspectives on the data in subsequent stages of analysis. In terms of the keyness analysis, WMatrix is one of the few tools that allows for such a detailed analysis based on a custom reference corpus, in addition to illustrating the resulting information in a variety of ways.

Part of Speech	Total No. of Lexemes	Freq. >10
Noun	2662	200
Verb	2581	107
Adjective	1197	128

Table 3: WMatrix words by part-of-speech

The total number of different nouns, verbs, and adjectives in the first book of *The Faerie Queene* is sizeable, with over 3000 nouns and verbs and over 1100 adjectives. However, when examining the frequencies of each of these words, the number that appear over ten times in the text is significantly lower: 199 nouns, 107 verbs, and 128 adjectives; the first twenty examples from each of the parts of speech are illustrated alongside their raw frequencies in Table 4. It is better to focus on words that occur relatively frequency in the text as the conclusions drawn about their significance are more representative of the text as a whole.

With regards to the CLAWS tagset, it was possible to establish the frequency of individual tags within the dataset. The benefits of this include a deeper understanding into the function of Spenser's word choice beyond the sometimes oversimplified categories of "verb" and "noun", although these fine distinctions within parts of speech resulted in minor discrepancies in total frequencies when compared across programs (see chapter 6). The results of the in-depth part-of-speech analysis were very simple for nouns and adjectives: the most common types of nouns were singular common (160 instances) and plural common (28 instances), with the

majority of the remaining tags relating to temporal or proper nouns. General adjectives were the most common type of adjectives with 121 instances, with only five superlatives and two comparatives featured in this dataset. The verb tags were more diverse, however. The most frequent types mostly focus on the past tense, with "past tense of lexical verb" and "past participle of lexical verb" featuring in the top three tags with thirty-two and eleven instances respectively. The second most frequent category was infinitives, with twenty-four instances, and the rest of the list consists of modal auxiliaries and forms of 'to have' and 'to be'.

The words with the highest keyness rating as identified by WMatrix are the most clearly illustrated in the word cloud below. In this representation, the larger a word is the higher its LL value (and therefore, its importance) is. In total, there are 324 words that appear in *The Faerie Queene* dataset more than ten times that have a statistically significant keyness value. Of these words, seventeen are not found in the EEBO reference corpus at all, and a further eighty-nine are found fewer than ten times.

Noun	Raw Freq.	Verb	Raw Freq.	Adjective	Raw Freq.
knight	169	*did*	489	*great*	125
life	72	*was*	326	*fair*	128
heart	69	*had*	149	*full*	70
man	66	*is*	148	*high*	66
day	62	*be*	147	*sad*	58
way	59	*began*	114	*false*	43
blood	57	*were*	99	*old*	42
death	55	*doth*	95	*goodly*	41
lady	55	*would*	88	*good*	40
hand	48	*could*	84	*dear*	37
quoth	46	*do*	71	*new*	37
eyes	46	*might*	64	*cruel*	36
night	45	*can*	61	*deadly*	35
foe	43	*said*	59	*living*	35
heaven	43	*does*	58	*other*	35
sight	43	*have*	58	*gentle*	34
Una	43	*hath*	57	*fierce*	33
ground	42	*may*	52	*heavenly*	33
place	40	*saw*	48	*true*	33
shield	39	*made*	46	*proud*	33

Table 4: Top twenty nouns, verbs, and adjectives

One observation that can be made by examining the total number of lexemes in this *Faerie Queene* dataset is that overall, the number of verbs and nouns is almost the same. However, when focusing on those that occur more than ten times, the number of verbs is half that of nouns. This is not completely surprising, given that recurring ideas and motifs require the reuse of the same nouns more so than verbs. Furthermore, lexical verbs are more likely to vary in order to add interest to a text, with a large portion of the recurring verbs falling under the category of auxiliary or modal verbs.

In examining the closed-class words, the emphasis on third person pronouns suggests that the majority of texts in the reference corpus are not written in the third person, setting *The Faerie Queene* apart from these other contemporary works. Upon examination of the results of the keyness analysis of open-class words, some of the flagged keywords are expected given the subject matter of the text. The importance of "knight" and "fairy", as well as characters like Duessa and Una, is predictable given their integrity to the text, yet there are other keywords that are more noteworthy. Flagging adjectives like "wondrous" and "feeble" is indicative of a level of descriptiveness present in *The Faerie Queene* that is not present in the EEBO reference corpus and hints at the importance of the reader in identifying relationships between words and senses. The difference between text- and reader-centric categorisations is a distinction worth making in corpus linguistics, and this will be discussed further in chapter 6.

Upon examination of the most frequent nouns as identified in WMatrix, it is to be expected that "knight" is at the top of the list; the first book of *The Faerie Queene* tells the tale of the Redcrosse Knight, who encounters yet more knights on his adventures. The theme of chivalry is also evidenced in other frequently used nouns like "foe", "shield", and "lady". Another important aspect of *The Faerie Queene* is the focus on religion, which is hinted at with the presence of "heaven" in the noun list, but there is greater evidence for this theme in other parts of speech.

A curious observation about the frequent noun list is the contrast between positive and negative words. "Blood", "death", "life" and "day" are all present in the top twenty most common nouns in this dataset, which indicates the polarisation of ideas in the text and possibly relates to the notion of good versus evil, a common feature of chivalric texts. The top of the

list, however, consists of words like “life” and “heart”, suggesting that the ultimate emphasis is on the triumph of positivity over negativity, which is consistent with the story’s plot.

The vast majority of the most frequent nouns are singular or plural common nouns, yet the character Una appears to be the only proper noun to appear on the list. This is not entirely accurate, since the Redcrosse Knight is counted under “knight”, but nevertheless, Una is the only female character to be present in the twenty most frequent nouns; Duessa, the second main female character, is thirty-fifth in the list. These high frequencies represent Una and Duessa’s importance within the story, as well as the significant role played by women throughout *The Faerie Queene*.

The list of the most frequently used verbs is filled with auxiliary and modal verbs which were not filtered out within WMatrix. However, a lot of these are past tense forms of verbs, which indicates the tense of the first book of *The Faerie Queene*. This correlates with the fact that Spenser is telling a story which supposedly took place a long time ago and adds to the mood of the story.

Aside from the modal and auxiliary verbs, the few lexical verbs present in the top twenty list – “began”, “said”, “saw”, and “made” – all relate to the active participation of the characters in the story. Such choices in verbs demonstrate that Spenser was more concerned with the actions of the characters than their thoughts and feelings, which corresponds with conventions of medieval and pre-16th-century literature, leaving their motives more open to interpretation by the reader.

The list of the most common adjectives corroborates the use of terms relating to chivalry, with examples like “great”, “fair”, and “true”, much in the same way as the list of nouns. There is also further evidence of the theme of good versus evil in the prevalence of “false”, “good”, “deadly”, and “true”. However, the frequent use of “goodly” and “heavenly” suggests a religious aspect to the text, with the prevalence of words such as “good” and “true” in the frequency lists indicating a Protestant focus (Smith, J.J. 2020, p.101). The significance of religion in *The Faerie Queene* has been much discussed (Hume 1984; King, J.N. 1990; Mallette 1997; Hadfield 2011), and the high frequency of these words reflects the importance of this theme beyond overt references.

In summation, the lists of the most frequent words in the first book of *The Faerie Queene* provides some interesting insights into the text. Some of these insights were to be expected; examples from the group of nouns and adjectives, as well as the most important keywords, reflect some of the most salient themes in the text, like chivalry, religion, and good versus evil. Yet the list of verbs showed that the few lexical verbs present suggested the focus of the story was on the characters' actions as opposed to what was going on inside their heads. Being able to determine this from a word list effectively illustrates the benefits of this method of analysis and provides the groundwork for subsequent stages of deeper analysis.

4.3 AntConc

4.3.1 Data

The process of gathering collocation data from AntConc was more time-consuming than the searches performed in WMatrix due to the nature of collocation searches and how AntConc deals with them. Gathering this data was useful for providing context for the words present in the frequency lists, thus bridging between individual words and the semantic categories of subsequent analyses. As before, the normalised text file of *The Faerie Queene* dataset was input into the program and the search parameters were set. In the "Collocation" tab, it is possible to choose the minimum frequency of collocates displayed, the distance to the left and right from the chosen word that the collocates are found, and by which parameter to sort the resulting collocate list. For this research, a minimum frequency of two narrowed the results to feasibly sized lists while still including as many potentially interesting collocates as possible. The default range of collocates in AntConc is within five words to the left or right and I saw no reason to change this. Furthermore, sorting the collocates by a statistical measure was the best way to highlight the most interesting collocates at the top of each list. AntConc allows for the use of different statistical measures, however using the default Mutual Information (MI) provided a sufficiently useful ranking system for this project (Stubbs 1995).

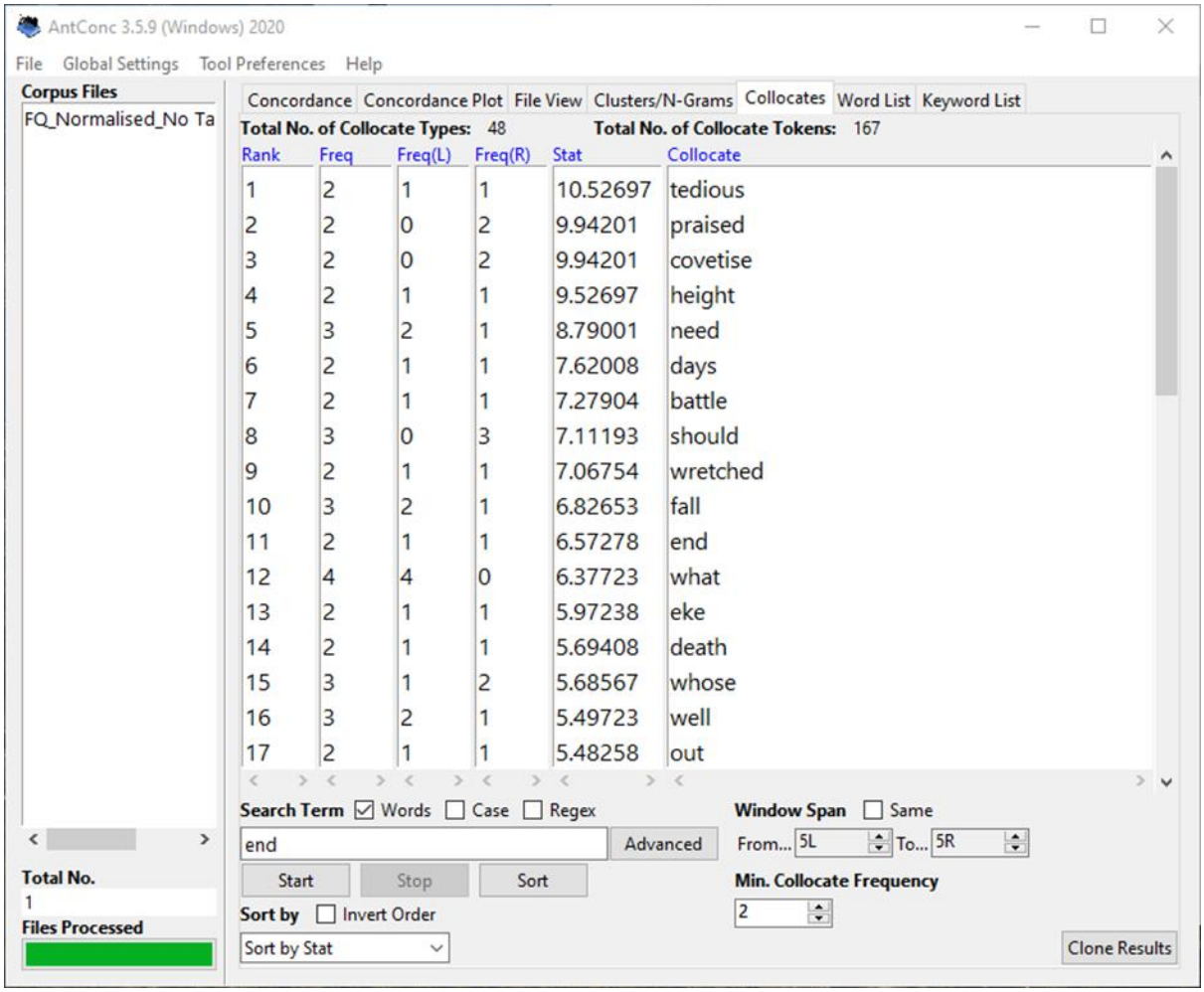


Figure 4: Collocation tab in AntConc

The WMatrix frequency lists formed the basis for the search terms; each noun, verb, and adjective that occur in the dataset more than ten times was input into AntConc and the results were subsequently cloned and copied into Excel tables. This step was necessary as AntConc does not allow you to save copies of results within the program and Excel allows for further filtering and refining of results. Another table was created within Excel detailing the total number of collocates for each search term, as well as notes of any interesting initial observations worthy of further consideration.

This particular methodology was best suited for this project because of its breadth and comprehensiveness. The choice to keep the AntConc search parameters as broad as possible, while avoiding creating an overwhelming amount of data, was made to avoid discounting elements that could be relevant to later stages of research. Nonetheless, this method resulted in a large body of data, which was made manageable by the filtering tools within Excel. Stacking

filters is not possible within AntConc, however being able to sort by MI then frequency in Excel gave the most statistically significant output with high frequencies.

The resulting data consisted of a table of collocates for each of the most frequent nouns, verbs, and adjectives in *The Faerie Queene* dataset, as well as a list of the lexemes with the highest number of collocates. The noun with the most collocates was "knight", with 211 distinct collocates. This is not unexpected given the subject matter of the text and the character the Redcrosse Knight. In a similar vein, the adjective with the highest number of collocates, 167, was "fair". The verb list was unusual, in that the verb with the largest number of collocates ("did", with 635 collocates) had approximately three times as many collocates as the second-place verb. Because of the huge number of collocates and the amount of work it would have entailed, "did" was discounted from analysis, leaving "is" as the verb with the highest number of collocates that was actually analysed. An exploration of the specific uses of "did" would be useful in future work, given that Early Modern English treats auxiliary verbs differently to present-day English, but this was beyond the scope of this project.

Not all of the lexemes from the WMatrix tables produced effective collocation tables. Due to not tagging this corpus to use in AntConc, some terms that exist as multiple parts of speech had all usages blended into one, resulting in strange hybrid lists. "Will" is a clear example of this, with both the noun and verb included in the same table. While in other cases this would not be impossible to overcome, like with "love", the ubiquity of "was" as a modal verb resulted in 414 collocates- too many to sort through to make differentiating the senses feasible.

In order to more easily facilitate the analysis of this data, lexemes were grouped into grammatical and semantic categories that were identified as being potentially interesting. These groups were:

- Singular and plural nouns
- Opposites
- Colours and materials
- Emotions
- Personal Names
- Near-synonyms

- Religion
- Chivalry and courtly traditions
- Violence and Battle
- Human Body
- Women

It is important to note that these themes are not based on semantic categories, but words were manually assigned to appropriate themes. Thematic associations are more relevant at this stage in the research because they allow for connections between semantically unrelated concepts that are nonetheless linked to a broader theme in the minds of people. These groups acted as a prelude to the subsequent semantic category analysis, yet they also allowed for an examination of how similar or related words are treated by Spenser at the lexical level.

4.3.2 Initial Observations

What becomes clear upon closer examination of the AntConc collocation data is that there are very few interesting (or open-class words) with a frequency higher than two. Keeping the minimum frequency so low was a deliberate choice in order to gather the most data, but the low frequency of noteworthy collocates emphasises the importance of MI as a measure for identifying the most relevant words. This overall low frequency is to be expected from a corpus of this size and subject material, as a larger corpus or one that includes more repetitive phrases would result in much higher frequency collocates. Observations can still be made using the available data, but conclusions can only be taken so far without a larger corpus to corroborate them.

A clear starting point in examining this data is with the three main characters of the story: the Redcrosse Knight, Una, and Duessa. In order to examine the collocates for “Redcrosse Knight” as a separate term from “knight”, the search parameters for “knight” were altered to include the context word “redcrosse”. The results indicated a distinct difference between how the Redcrosse Knight and knights generally are treated in the text; some of the most frequent collocates of “knight” are “prowest”, “errant”, and “doughty”, whereas with “redcrosse knight”, “gentle”, “love”, and “good” appear at the top of the list. While both groups of collocates refer to ideas associated with chivalry (Friedman 2019, p.91), the Redcrosse Knight

appears to embody different ideals to that of other knights in the text. A further examination of the concordances will establish whether this is an accurate assessment or not.

The two main female characters in the first book of *The Faerie Queene* are Una and Duessa. Both women have very different roles within the plot, and this is demonstrated in their collocate lists. The top collocates for Una include “oppressed”, “fairest”, “besought”, and “woeful”, placing her as a damsel-in-distress figure. This is a classic stereotype within chivalric literature (Altermatt 2001, p.11), which correlates with how Una is treated in the first book of *The Faerie Queene.* By contrast, Duessa’s top collocates include “false”, “foul”, and “proud”. In contrast with other relevant collocates having a low frequency, “false” appears in close proximity with “Duessa” seventeen times. Such a high frequency emphasises her duplicitousness and reinforces her role as a villainous character. However, the collocate with the highest MI is “wept”, which could be considered unusual. No other character appears on the collocate list of “wept” save Duessa, suggesting a notable lack of composure on her part. Combined with her other negative attributes, this could mean that such an outward display of negative emotion is undesirable. This idea will be explored further in the Discussion section.

Examining the treatment of these three main characters and their collocates results in some interesting observations. However, it is clear that each of the characters is in their own way a stereotype of some aspect of chivalric literature. Whether good, evil, heroic, or a fair maiden, Spenser uses his characters as representations of chivalric notions rather than real people, indicating that their purpose is symbolic as opposed to a genuine representation of humanity.

As previously stated, the words included in the group “Chivalry and Courtly Traditions” would not necessarily be found in this semantic category but are associated with this genre and related themes. However, the collocates were not factored into the judgements made about these words. It is therefore worth considering how many of the terms in this group have clear and tangible ties to the theme of chivalry. Most of the terms with interesting open-class collocates were relevant in some way to the theme, but a few words typically associated with chivalry were notable devoid of pertinent collocates: “knights”, “dwarf”, and “glory”. The lack of collocates for “knights” is counterbalanced by the considerable number for “knight”, indicating that the plural form is less relevant to the story than the singular. Considering the text’s plot and structure, this makes sense, as most of the story deals with only the Redcrosse

Knight, with few encounters with other knights or general statements about knights. The dwarf is a recurring figure in chivalric literature whose function is to "act as symbols of their respective knights' aggression, driving the knights into the adventure of the plot" (Carroll 2018, p.48). Given his important function, it is noteworthy that the dwarf in the first book of *The Faerie Queene* has no relevant collocates, but this can be explained by the fleetingness of the character within the story as a whole. Lastly, the lack of collocates for "glory" hints at there being other motivations or chivalric ideals that are more important in the story than personal acclaim.

Of the words whose collocates confirm ties to the theme of chivalry, there are some that describe positive chivalric attributes. For "courage", the top collocates include "zeal", "youthly", "sturdy", "bold", and "stout", indicating a kind of masculine bravery is most desirable in the context of the story. This fits with the plot of *The Faerie Queene*, where knights have the most agency and strength in battle is one of the most valuable qualities a male character can have (Friedman 2019, p.89). "King" is in a similar position, with "mighty" and "proud" appearing high in the collocates list, emphasising the importance of power and strength in ruling. By contrast, the most significant collocates for "squire" include "loved", "gentle", and "mild", which is similar to the attributes found in the collocate list of the Redcrosse Knight. This possibly relates to the Redcrosse Knight being someone squires aspired to be like and the qualities required to do so, but a more detailed examination of the concordances is needed to confirm this.

In opposition to the words and collocates describing positive attributes, others describe negative characteristics. Given that a dragon – whether a literal dragon or other such danger – is a standard chivalric adversary for good and noble knights, it is not surprising that its top collocates are "horrible", "stretched", and "cursed". Such associations only reinforce the evil nature of dragons in chivalric literature. However, the most significant collocates of "proud" suggest that the negative connotations are derived from religion: "pagan", "Lucifera", and "Saracen". In a Protestant setting, illustrating the malicious and immoral connotations of one of the seven deadly sins using religious terms is an effective way of reinforcing the idea of good versus evil within a chivalric context.

During the process of gathering the collocation data from AntConc, it became increasingly clear that there were many of words pertaining to the human body present. It therefore made sense to examine these terms as a group to see if there was anything noteworthy in this data.

As is to be expected, a number of words in this group were literal body parts with their collocates being descriptive terms. With “breast”, the most significant collocates were “riven”, “breath”, and “bloody”, clearly indicating that the correct sense of “breast” in this context is a person’s chest. The violent connotations of the collocates also suggest that the term in used more in masculine contexts as opposed to feminine; battles in chivalric literature are almost always fought by men, with women either watching from the sidelines or not involved at all. Another interesting word with descriptive collocates is “flesh”, whose top collocates include “impair”, “frail”, and “die”. Such fallible connotations for these collocates hint at “flesh” being a negative counterpart to another word such as “body”. The most significant collocates for “body” include “fit”, “dead”, and “sore”, which – while not necessarily being positive – suggest a robustness not displayed with “flesh”. An examination of the concordances of these collocates will determine whether this analysis is accurate.

Some of the other body parts listed in this group were not used literally, but rather as a representation of other ideas. “Heart” is an obvious example of this, with its top collocates being “sighing”, “melt”, “feel”, and “anguish”. While illustrating a variety of emotions, it is clear from these collocates that “heart” refers more to the emotional core within people rather than the literal organ. A similar pattern can be found with “tongue” and its collocates “faltering” and “wit”. Here the tongue represents the concept of speech which – much like heart – is a very common sense and application of the term. While not a new revelation in and of itself, it is beneficial to see that these nuances in meaning and usage are clear from a word’s list of collocates.

4.4 LIMITATIONS

The lexical frequency and collocation portion of this research was completed successfully and resulted in a large amount of data, but there were limitations that must be acknowledged. One of the main challenges in using AntConc was not providing it with a part-of-speech tagged text, which would have been beneficial in distinguishing between the noun and verb forms of words

like “will” and “love”. Furthermore, this research was completed without using a stoplist, a list of words that is input into the software and removed from searches. This feature would have helped remove the extraneous words that made the AntConc data laborious to work through. Increased familiarity with the software and methodology means that in future research this feature will be employed.

5 Core Semantic Categories: Semantic Analysis

This chapter describes the process by which the first book of *The Faerie Queene* was semantically tagged according to the taxonomy of the Historical Thesaurus of English. First, the organisation of the HT and the inner workings of the semantic tagger are detailed, leading to an exploration of the resulting data. Then, the data visualisation tool Treemap is discussed, along with the information that such representations can provide. The chapter concludes with a description of the limitations of using these resources.

5.1 About the Historical Thesaurus

The Historical Thesaurus of English (HT) is a comprehensive thesaurus of English words spanning the entirety of the language's history and is the first thesaurus of this depth for any language. The project was started in 1965 by Michael Samuels at the University of Glasgow and the first edition, published in October 2009, was the result of decades of painstaking work undertaken by many researchers and project assistants. The thesaurus exists as a two-volume print edition and an online resource that is regularly updated.[5] The second edition of the Historical Thesaurus was published in October 2020 and features updated lexical data and category revisions.[6] In this research, the first edition of the Historical Thesaurus will be used because some of the tools used have not been updated to match the category hierarchy of the second edition.

The data for the Historical Thesaurus is sources from the Oxford English Dictionary (OED) and, for words dating from c.750 to 1100, A Thesaurus of Old English. As the OED add new senses and update their entries, changes are made to the Historical Thesaurus accordingly. The thesaurus is organised into three major sections: "The World", "The Mind", and "Society". Each of these sections are then subdivided into major categories like "Emotion", "Movement", and "Leisure", with further subdivisions describing ever more specific concepts. Ultimately, the words within each category are organised according to part of speech and then ordered chronologically. The numbers for each category heading reflect this hierarchical structure:

[5] Available at https://ht.ac.uk/
[6] A full list of revisions is available at https://ht.ac.uk/versions-and-changes/

01 The World

→ 01.04 People

→ 01.04.04 Person

→ 01.04.04.04 Child

→ 01.04.04.04.02 Girl

This numbering system means that the position of a particular category within the hierarchy of the Historical Thesaurus is easy to determine, as well as establishing hyper- and hyponyms of words within categories.

5.2 About the Semantic Tagger

The Historical Thesaurus Semantic Tagger (HTST) is a product of the SAMUELS Project (Semantic Annotation and Mark-Up for Enhancing Lexical Searches) alongside the Semantic Hansard Corpus and Semantic EEBO. The project is a collaboration of researchers from the University of Glasgow and the University of Lancaster's University Centre for Computer Corpus Research on Language (UCREL), including the creators of VARD and the USAS Tagger. The purpose of the HTST is to tag each word in the inputted texts with the most accurate category

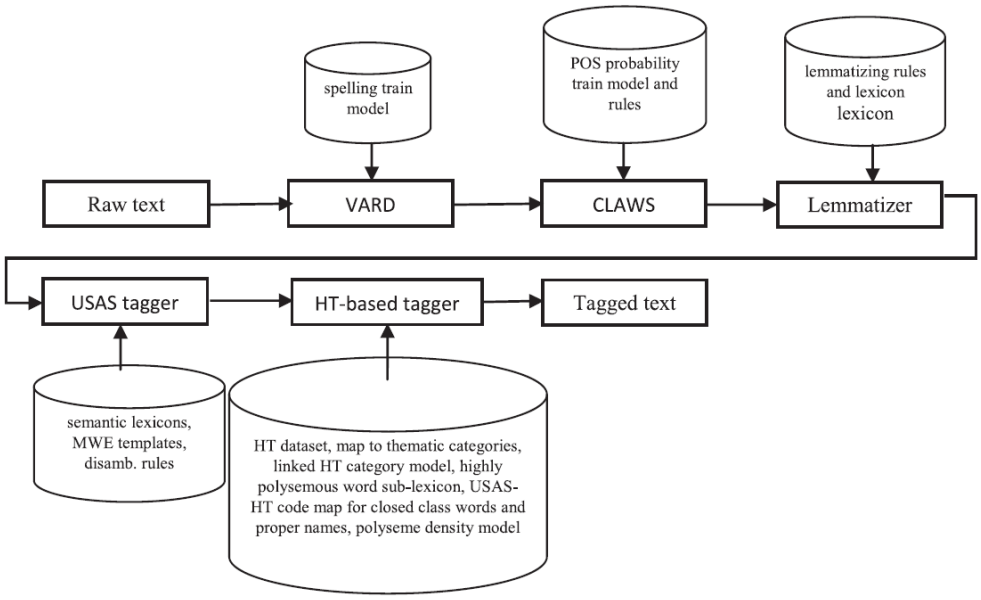
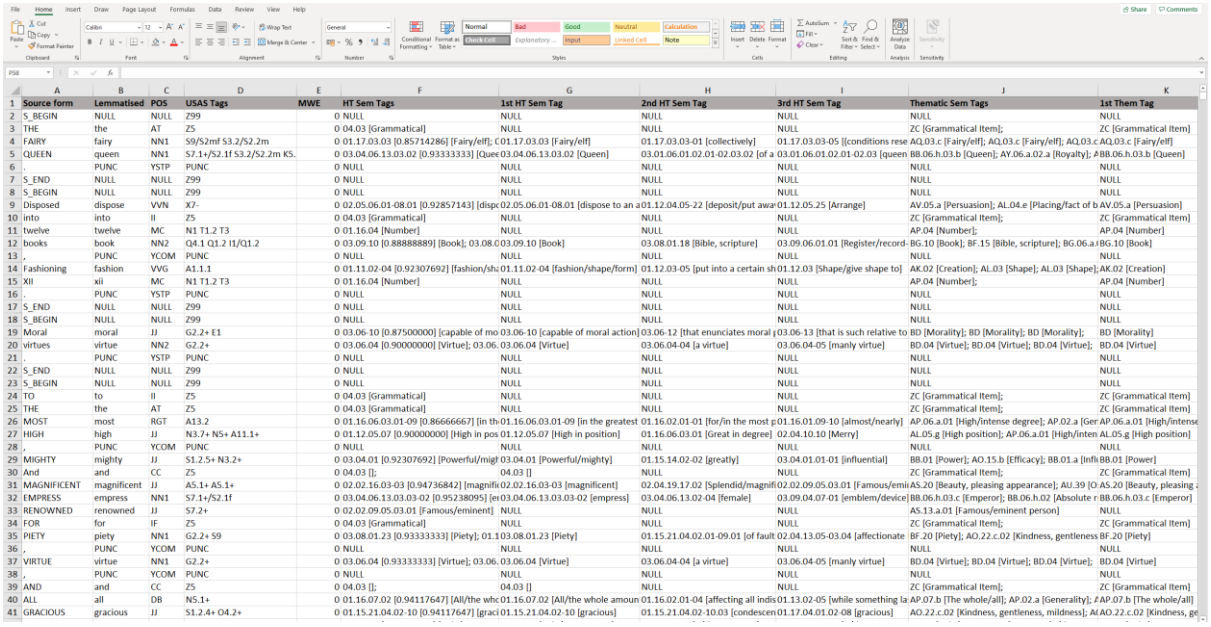


Figure 5: Architecture of the Historical Thesaurus Semantic Tagger system (Piao et al. 2017, p.117)

in the Historical Thesaurus (HT), with the resulting data having a myriad of further uses across multiple linguistic fields.

The tagger contains a copy of VARD (see chapter four), the CLAWS part-of-speech tagger, and the USAS semantic tagger; the text is run through VARD first to ensure regularity of spelling, then the CLAWS and USAS tagger to identify the parts of speech and provide a baseline for semantic categorisation. Once those stages are completed, the text is run through a tagger based on a streamlined categorisation structure of the HT. The changes made to the Historical Thesaurus structure used in the semantics tagger include reducing the number of fine-grained categories to minimise confusion and adding grammatical categories for closed-class words like pronouns and conjunctions. In order to determine the correct semantic category for each word, the tagger uses data from the HT and the results from the USAS tagger.[7] The results from all of these tools are then output into a spreadsheet for further use.



	A	B	C	D	E	F	G	H	I	J	K
1	Source form	Lemmatised	POS	USAS Tags	MWE	HT Sem Tags	1st HT Sem Tag	2nd HT Sem Tag	3rd HT Sem Tag	Thematic Sem Tags	1st Them Tag
2	S_BEGIN	NULL	NULL	Z99	0	NULL	NULL	NULL	NULL	NULL	NULL
3	THE	the	AT	Z5	0	04.03 [Grammatical]	NULL	NULL	NULL	ZC [Grammatical Item];	ZC [Grammatical Item]
4	FAIRY	fairy	NN1	S9/S2mf S3.2/S2.2m	0	01.17.03.03 [0.85714286] [Fairy/elf]; C	01.17.03.03 [Fairy/elf]	01.17.03.03-01 [collectively]	01.17.03.03-05 [(conditions rese	AQ.03.c [Fairy/elf]; AQ.03.c [Fairy/elf]; AQ.03.c	AQ.03.c [Fairy/elf]
5	QUEEN	queen	NN1	S7.1+/S2.1f S3.2/S2.2m K5.	0	03.04.06.13.03.02 [0.93333333] [Quee	03.04.06.13.03.02 [Queen]	03.01.06.01.02.01-02.03.02 [of a	03.01.06.01.02.01-02.03 [queen	BB.06.h.03.b [Queen]; AY.06.a.02.a [Royalty]; A	BB.06.h.03.b [Queen]
6	.	PUNC	YSTP	PUNC	0	NULL	NULL	NULL	NULL	NULL	NULL
7	S_END	NULL	NULL	Z99	0	NULL	NULL	NULL	NULL	NULL	NULL
8	S_BEGIN	NULL	NULL	Z99	0	NULL	NULL	NULL	NULL	NULL	NULL
9	Disposed	dispose	VVN	X7-	0	02.05.06.01-08.01 [0.92857143] [dispo	02.05.06.01-08.01 [dispose to an a	01.12.04.05-22 [deposit/put awa	01.12.05.25 [Arrange]	AV.05.a [Persuasion]; AL.04.e [Placing/fact of b	AV.05.a [Persuasion]
10	into	into	II	Z5	0	04.03 [Grammatical]	NULL	NULL	NULL	ZC [Grammatical Item];	ZC [Grammatical Item]
11	twelve	twelve	MC	N1 T1.2 T3	0	01.16.04 [Number]	NULL	NULL	NULL	AP.04 [Number];	AP.04 [Number]
12	books	book	NN2	Q4.1 Q1.2 I1/Q1.2	0	03.09.10 [0.88888889] [Book]; 03.08.0	03.09.10 [Book]	03.08.01.18 [Bible, scripture]	03.09.06.01.01 [Register/record-	BG.10 [Book]; BF.15 [Bible, scripture]; BG.06.a	BG.10 [Book]
13	,	PUNC	YCOM	PUNC	0	NULL	NULL	NULL	NULL	NULL	NULL
14	Fashioning	fashion	VVG	A1.1.1	0	01.11.02-04 [0.92307692] [fashion/sha	01.11.02-04 [fashion/shape/form]	01.12.03-05 [put into a certain sh	01.12.03 [Shape/give shape to]	AK.02 [Creation]; AL.03 [Shape]; AL.03 [Shape];	AK.02 [Creation]
15	XII	xii	MC	N1 T1.2 T3	0	01.16.04 [Number]	NULL	NULL	NULL	AP.04 [Number];	AP.04 [Number]
16	.	PUNC	YSTP	PUNC	0	NULL	NULL	NULL	NULL	NULL	NULL
17	S_END	NULL	NULL	Z99	0	NULL	NULL	NULL	NULL	NULL	NULL
18	S_BEGIN	NULL	NULL	Z99	0	NULL	NULL	NULL	NULL	NULL	NULL
19	Moral	moral	JJ	G2.2+ E1	0	03.06-10 [0.87500000] [capable of mo	03.06-10 [capable of moral action]	03.06-12 [that enunciates moral p	03.06-13 [that is such relative to	BD [Morality]; BD [Morality]; BD [Morality];	BD [Morality]
20	virtues	virtue	NN2	G2.2+	0	03.06.04 [0.90000000] [Virtue]; 03.06.	03.06.04 [Virtue]	03.06.04-04 [a virtue]	03.06.04-05 [manly virtue]	BD.04 [Virtue]; BD.04 [Virtue]; BD.04 [Virtue];	BD.04 [Virtue]
21	.	PUNC	YSTP	PUNC	0	NULL	NULL	NULL	NULL	NULL	NULL
22	S_END	NULL	NULL	Z99	0	NULL	NULL	NULL	NULL	NULL	NULL
23	S_BEGIN	NULL	NULL	Z99	0	NULL	NULL	NULL	NULL	NULL	NULL
24	TO	to	II	Z5	0	04.03 [Grammatical]	NULL	NULL	NULL	ZC [Grammatical Item];	ZC [Grammatical Item]
25	THE	the	AT	Z5	0	04.03 [Grammatical]	NULL	NULL	NULL	ZC [Grammatical Item];	ZC [Grammatical Item]
26	MOST	most	RGT	A13.2	0	01.16.06.03.01-09 [0.86666667] [in th	01.16.06.03.01-09 [in the greatest	01.16.02.01-01 [for/in the most p	01.16.01.09-10 [almost/nearly]	AP.06.a.01 [High/intense degree]; AP.02.a [Ger	AP.06.a.01 [High/intense
27	HIGH	high	JJ	N3.7+ N5+ A11.1+	0	01.12.05.07 [0.90000000] [High in pos	01.12.05.07 [High in position]	01.16.06.03.01 [Great in degree]	02.04.10.10 [Merry]	AL.05.g [High position]; AP.06.a.01 [High/inten	AL.05.g [High position]
28	,	PUNC	YCOM	PUNC	0	NULL	NULL	NULL	NULL	NULL	NULL
29	MIGHTY	mighty	JJ	S1.2.5+ N3.2+	0	03.04.01 [0.92307692] [Powerful/migh	03.04.01 [Powerful/mighty]	01.15.14.02-02 [greatly]	03.04.01.01-01 [influential]	BB.01 [Power]; AO.15.b [Efficacy]; BB.01.a [Infl	BB.01 [Power]
30	And	and	CC	Z5	0	04.03 [];	04.03 []	NULL	NULL	ZC [Grammatical Item];	ZC [Grammatical Item]
31	MAGNIFICENT	magnificent	JJ	A5.1+ A5.1+	0	02.02.16.03-03 [0.94736842] [magnifi	02.02.16.03-03 [magnificent]	02.04.19.17.02 [Splendid/magnifi	02.02.09.05.03.01 [Famous/emi	AS.20 [Beauty, pleasing appearance]; AU.39 [O	AS.20 [Beauty, pleasing
32	EMPRESS	empress	NN1	S7.1+/S2.1f	0	03.04.06.13.03.03-02 [0.95238095] [e	03.04.06.13.03.03-02 [empress]	03.04.06.13.02-04 [female]	03.09.04.07-01 [emblem/device]	BB.06.h.03.c [Emperor]; BB.06.h.02 [Absolute r	BB.06.h.03.c [Emperor]
33	RENOWNED	renowned	JJ	S7.2+	0	02.02.09.05.03.01 [Famous/eminent]	NULL	NULL	NULL	AS.13.a.01 [Famous/eminent person]	NULL
34	FOR	for	IF	Z5	0	04.03 [Grammatical]	NULL	NULL	NULL	ZC [Grammatical Item];	ZC [Grammatical Item]
35	PIETY	piety	NN1	G2.2+ S9	0	03.08.01.23 [0.93333333] [Piety]; 01.1	03.08.01.23 [Piety]	01.15.21.04.02.01-09.01 [of fault	02.04.13.05-03.04 [affectionate	BF.20 [Piety]; AO.22.c.02 [Kindness, gentleness	BF.20 [Piety]
36	,	PUNC	YCOM	PUNC	0	NULL	NULL	NULL	NULL	NULL	NULL
37	VIRTUE	virtue	NN1	G2.2+	0	03.06.04 [0.93333333] [Virtue]; 03.06.	03.06.04 [Virtue]	03.06.04-04 [a virtue]	03.06.04-05 [manly virtue]	BD.04 [Virtue]; BD.04 [Virtue]; BD.04 [Virtue];	BD.04 [Virtue]
38	,	PUNC	YCOM	PUNC	0	NULL	NULL	NULL	NULL	NULL	NULL
39	AND	and	CC	Z5	0	04.03 [];	04.03 []	NULL	NULL	ZC [Grammatical Item];	ZC [Grammatical Item]
40	ALL	all	DB	N5.1+	0	01.16.07.02 [0.94117647] [All/the whc	01.16.07.02 [All/the whole amoun	01.16.02.01-04 [affecting all indis	01.13.02-05 [while something la	AP.07.b [The whole/all]; AP.02.a [Generality]; A	AP.07.b [The whole/all]
41	GRACIOUS	gracious	JJ	S1.2.4+ O4.2+	0	01.15.21.04.02-10 [0.94117647] [graci	01.15.21.04.02-10 [gracious]	01.15.21.04.02-10.03 [condescen	01.17.04.01.02-08 [gracious]	AO.22.c.02 [Kindness, gentleness, mildness]; A	AO.22.c.02 [Kindness, ge

Figure 6: Semantic Tagger output data

5.3 METHOD

In order to gather the initial data from the HTST, the version of the first book of *The Faerie Queene* with normalised spelling was input into the tagger. The problems encountered during

[7] For more detail about how the HTST works, see Piao et al. 2017.

this process are detailed in the Limitations section. Once completed, the tagger's output was the sizeable table of codes illustrated in part in Figure 6.

Given the size of the output, there were multiple stages to organising this data and making it more relevant and manageable. First, the PivotTables function in Excel was used to create a list of the most frequent Historical Thesaurus categories. This list is essential to further stages of analysis as it provides a clear representation of where the potentially significant aspects of *The Faerie Queene* category data lie, despite the most frequent categories being closed-class words.

The tagger outputs the three most likely tags for each word in order of the tagger's confidence in one spreadsheet column; for this project, these tags were split into their own separate columns. For example, the word "piety" is assigned to the following three categories, with the first being the most likely to be accurate (according to the tagger):

1. 03.08.01.23 [Piety]
2. 01.15.21.04.02.01-09.01 [of faults of others]
3. 02.04.13.05-03.04 [affectionate loyalty]

Not every word has three potential assignations, but there is still an overwhelming amount of data if all three columns are included in the analysis. Therefore, only the first column listing the most likely category for each word was included in the PivotTable.

Once the list of most frequent categories in *The Faerie Queene* dataset was completed, all categories with a frequency higher than twenty were copied into another table and broken down into their first-, second-, and third-level tier headings, as well as their parts of speech. For instance, category 01.09.02.03 [Weary/exhausted] is listed in the Historical Thesaurus as an adjectival category, and it can be determined from the category number that the first three tier headings are:

First tier: 01 The World
Second tier: 01.09 Physical Sensation
Third tier: 01.09.02 Sleeping and Waking

The decision was made to only include the first three tiers because most of the categories with high frequencies had at least three tiers to list; a small number had fewer, but not enough to warrant making the cut-off point at a higher tier. By recording the high-lever tiers of each category, it allows for different levels and scales of analysis; it is possible that there are wider thematic trends within the first book of *The Faerie Queene* that can only be uncovered by looking at more general categories. Furthermore, by including the part of speech of each category, the grammatical contexts of the listed categories can be considered in further stages of analysis.

5.4 Semantic Tagger Data

The data resulting from the HTST and subsequent reorganisations of the raw data culminated in two main sections: overall category frequency, and frequency of tier one, two, and three category headings. Starting with overall frequencies, the HTST identified 5454 distinct semantic categories within the first book of *The Faerie Queene*. However, 2583 of these categories – around 47% – are only found once. To reduce the number of categories to a manageable level, the cut-off minimum frequency of twenty was chosen, resulting in 217 categories. Of these categories, the top six most frequent describe closed-class grammatical words, with the seventh referring to the verb "be". Unsurprisingly, the most frequent category of note is 03.01.06.01.02.09 [Knight], with 174 occurrences.

There are some interesting shifts in importance that happen when these categories are divided into their tier one, two, and three categories. In tier one, once grammatical words are removed, "The World" is found to be the largest category, with 6976 occurrences out of 11,265. When included, grammatical terms take up 44% of tier one usages, so in order to gain a more representative view of lexical semantic categories, grammatical categories are discounted.

In tier two, "Possession", "Relative Properties", and "Time" have the lion's share of occurrences, which is understandable given their integrity to coherent language; "his" and "her" are among the most common words in "Possession," which corresponds with *The Faerie Queene* being written in third person; words like "all" and "each" are included in "Relative Properties"; and times of day, as well as adjectives like "long", make up the "Time" category.

Tier three tells a similar tale, with “Owning” and “Existence” topping the list with the highest frequencies. However, the third most frequent category is “The Body”, which corroborates the observation made about frequency of body part terms during the AntConc stage of research. In order to gain a better understanding of this data, however, it is beneficial to examine it in an alternative way to tables and statistics, by using visual representation tools.

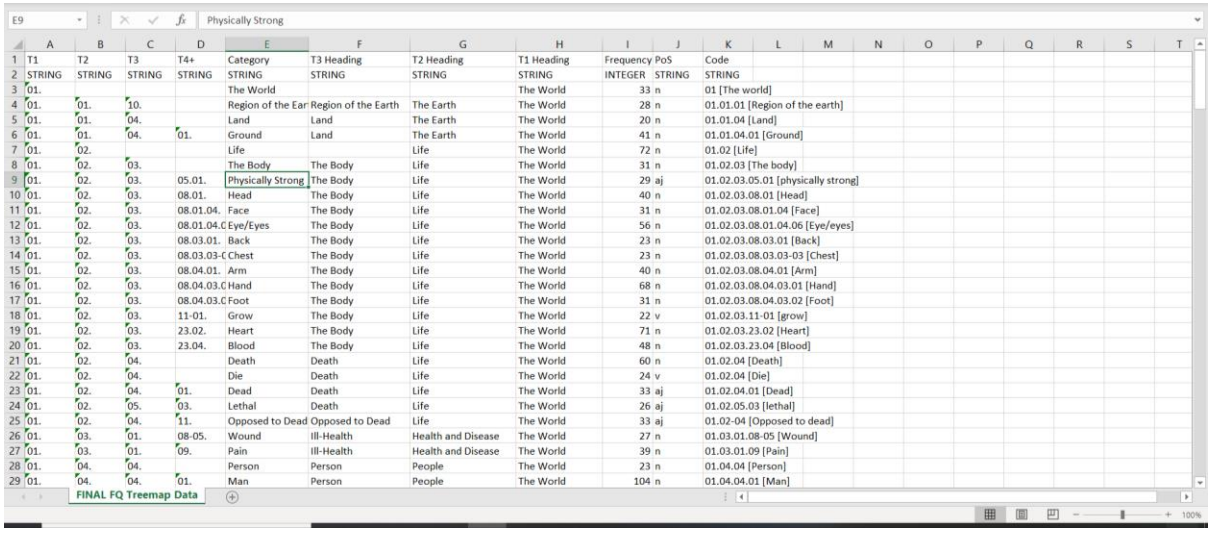
5.5 About Treemap

Treemap is a data visualisation program stemming from the Human-Computer Interaction Lab at the University of Maryland. It was first designed by Ben Shneiderman in the 90s and was updated by a number of others until the mid-2000s. It is described as a tool for providing a “space-constrained visualisation of hierarchical structures” (known as ‘treemaps’) that uses size and colour to provide more information about leaf nodes at various depths within the diagram (Plaisant 2003).

The version of the program used for this project is Treemap 4.1, which was released in 2004. Despite the age of the tool, it still provides a clear visual representation of hierarchical data- a useful way to interpret the data from the HST. By having a graphic depiction of the most frequent categories in *The Faerie Queene* dataset, it becomes easier to examine the size and significance of individual categories and tiers, as well as compare them to the size of those found in the HT proper.

5.6 Treemap Methodology

In order to create the treemaps, the data collected from the HTST needed to be converted to the correct format for the program. Treemap’s preferred input is .tm3 files, and the best way to create these is to create a tab-delimited .txt file and manually rename the file to the correct type. Excel was used to create these .txt files as it facilitated easy copying from the original dataset. The data required to create the best treemap was category headings; the names of



T1	T2	T3	T4+	Category	T3 Heading	T2 Heading	T1 Heading	Frequency	PoS	Code
STRING	STRING	STRING	STRING	STRING	STRING	STRING	STRING	INTEGER	STRING	STRING
01.				The World			The World	33	n	01 [The world]
01.	01.	10.		Region of the Ear	Region of the Earth	The Earth	The World	28	n	01.01.01 [Region of the earth]
01.	01.	04.		Land	Land	The Earth	The World	20	n	01.01.04 [Land]
01.	01.	04.	01.	Ground	Land	The Earth	The World	41	n	01.01.04.01 [Ground]
01.	02.			Life		Life	The World	72	n	01.02 [Life]
01.	02.	03.		The Body	The Body	Life	The World	31	n	01.02.03 [The body]
01.	02.	03.	05.01.	Physically Strong	The Body	Life	The World	29	aj	01.02.03.05.01 [physically strong]
01.	02.	03.	08.01.	Head	The Body	Life	The World	40	n	01.02.03.08.01 [Head]
01.	02.	03.	08.01.04.	Face	The Body	Life	The World	31	n	01.02.03.08.01.04 [Face]
01.	02.	03.	08.01.04.C	Eye/Eyes	The Body	Life	The World	56	n	01.02.03.08.01.04.06 [Eye/eyes]
01.	02.	03.	08.03.01.	Back	The Body	Life	The World	23	n	01.02.03.08.03.01 [Back]
01.	02.	03.	08.03.03-C	Chest	The Body	Life	The World	23	n	01.02.03.08.03.03-03 [Chest]
01.	02.	03.	08.04.01.	Arm	The Body	Life	The World	40	n	01.02.03.08.04.01 [Arm]
01.	02.	03.	08.04.03.C	Hand	The Body	Life	The World	68	n	01.02.03.08.04.03.01 [Hand]
01.	02.	03.	08.04.03.C	Foot	The Body	Life	The World	31	n	01.02.03.08.04.03.02 [Foot]
01.	02.	03.	11-01.	Grow	The Body	Life	The World	22	v	01.02.03.11-01 [grow]
01.	02.	03.	23.02.	Heart	The Body	Life	The World	71	n	01.02.03.23.02 [Heart]
01.	02.	03.	23.04.	Blood	The Body	Life	The World	48	n	01.02.03.23.04 [Blood]
01.	02.	04.		Death	Death	Life	The World	60	n	01.02.04 [Death]
01.	02.	04.		Die	Death	Life	The World	24	v	01.02.04 [Die]
01.	02.	04.	01.	Dead	Death	Life	The World	33	aj	01.02.04.01 [Dead]
01.	02.	05.	03.	Lethal	Death	Life	The World	26	aj	01.02.05.03 [lethal]
01.	02.	04.	11.	Opposed to Dead	Opposed to Dead	Life	The World	33	aj	01.02-04 [Opposed to dead]
01.	03.	01.	08-05.	Wound	Ill-Health	Health and Disease	The World	27	n	01.03.01.08-05 [Wound]
01.	03.	01.	09.	Pain	Ill-Health	Health and Disease	The World	39	n	01.03.01.09 [Pain]
01.	04.	04.		Person	Person	People	The World	23	n	01.04.04 [Person]
01.	04.	04.	01.	Man	Person	People	The World	104	n	01.04.04.01 [Man]

Figure 7: Treemap input text file

tiers one, two, and three; part of speech; category frequency; and category numbers from the HT. I also included these category numbers broken down by tier, but this was not strictly necessary beyond ease of comprehension of the tables. The purpose of including the parts of speech was to facilitate the creation of separate tables for nouns, adjectives, and verbs without having to go through the entire copying process multiple times. Furthermore, including the category frequencies is essential for establishing the size of each category in the treemap. However, the sole purpose of including the entire HT category number at the end of each row was to sort the table alphabetically in order to check for errors.

The formatting requirements described in the Treemap documentation require each column to be labelled ‘string’ or ‘integer’, as seen in figure 7. It is important to include an integer column if the desired effect is to include size variation in the final treemap; the program cannot read a column labelled ‘string’ as a valid selection for changing node size.

Once the formatting is complete, the file is saved as a tab-delimited .txt file and manually renamed to a .tm3 file. Copies of both files were kept in case changes needed to be made to the original table, as converting the .tm3 file back to a .txt resulted in file corruption. This process was repeated for the categories labelled as nouns, adjectives, and verbs, in order to provide a finer level of detail for each of these parts of speech. After each .tm3 file was created, they could then be loaded into Treemap.

In order to provide an effective point of reference for comparing category size, a reference treemap was made using the Thematic Category Set for the HT.[8] This streamlined list of HT categories was the easiest way to create a useful point of reference that did not involve using the entire HT dataset or tagging every text used in the WMatrix reference corpus, which would have been an ideal point of comparison but was impractical within the limitations of this project. The Thematic Category Set is useful as a data source for a reference treemap because it acts as an effective indicator as to the size of categories within the full data of the HT, so it becomes possible to see how the size of categories within *The Faerie Queene* dataset differ from those of the English language as a whole. However, it means that the semantic distribution of *The Faerie Queene* cannot be compared directly with Early Modern English. The HT reference treemap was created by following the same process as *The Faerie Queene* treemaps- converting the Excel file into a correctly formatted .tm3 file and uploading to Treemap.

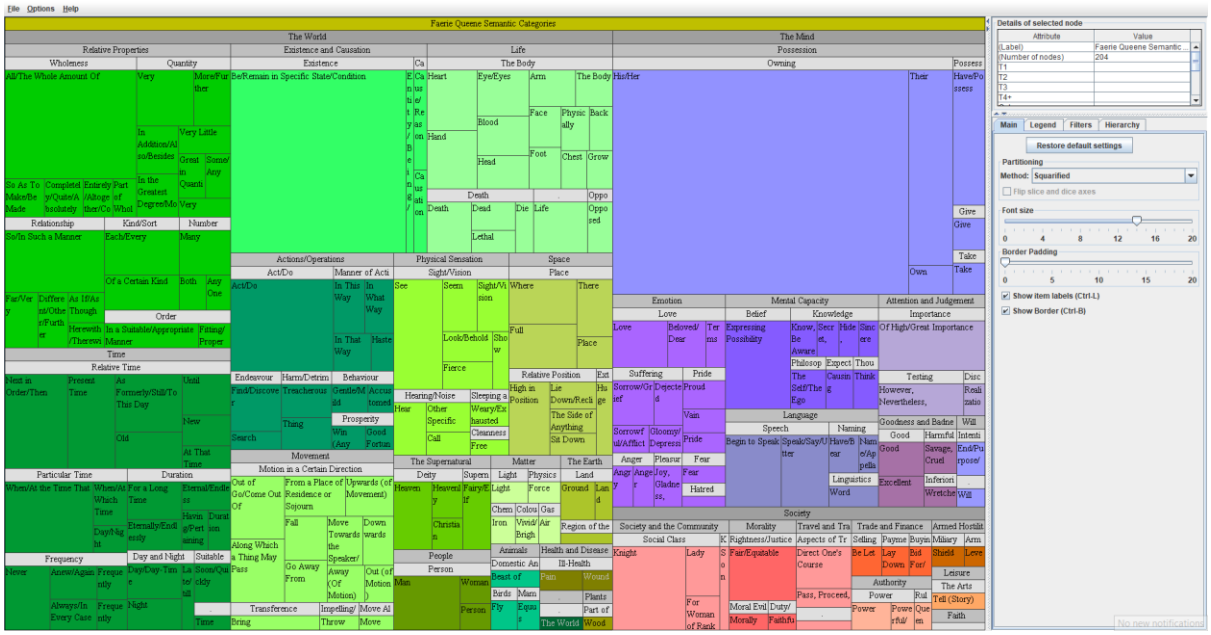


Figure 8: Treemap user interface

The user interface of Treemap features the representation of the uploaded data covering most of the screen, with a panel on the right-hand side showing details about a node, if selected, as

[8] Available at https://www.gla.ac.uk/schools/critical/research/fundedresearchprojects/samuels/outputs/#toolsandcorpora

well as four options tabs. “Main” allows you to choose the graphic option best suited to the data, with the options being “squarified” (as can be seen in figure 8), “slice and dice”, and “strip”. In this tab, you can also choose the font size of the labels and the size of the box borders, if any. “Legend” is where you set the parameters of the Treemap by choosing how specific columns from the inputted table are used. In this particular case, the box labels are based on the “Category” column, box size on “Frequency”, and colours chosen according to “T2 Heading”. This tab also allows you to change the colours used for each heading; categories under *“*The World*”* are coloured different shades of green, those under *“*The Mind*”* use purple, and *“*Society*”* use reds and oranges. These choices of colours and shades of categories located next to each other was made so that the tier one and two categories would be easily differentiated by those with colour blindness, and these selevctions was applied to each of the subsequent treemaps created. The choice to use the headings from tier 2 of the HT as the finest level of colour detail allowed for a clear representation of broader categories without requiring different colours for every category shown, as well as making the size of different tiers of categories clear from just a glance. The third tab labelled “Filters” is useful if you wish to filter out certain leaf nodes or sub-trees. The final tab, “Hierarchy” was necessary for determining the hierarchy of the tiers within the Treemap. It is also possible to set multiple hierarchies, but this was not necessary for this project.

Once the treemap illustrating all the semantic categories from *The Faerie Queene* dataset was completed, it was saved as a .tms file (human-readable text file of settings) and a .png for more general applications. This entire process was repeated for the tables of noun, verb, and adjective categories, as well as the Thematic Category reference set, since Treemap does not allow for the automatic application of settings from one file to another.

5.7 Initial Observations

5.7.1 Semantic Tagger

When looking at the frequencies of specific categories, it is to be expected that the categories that top the list are grammatical words and pronouns. The construction of coherent sentences requires an abundance of these words, so the very high frequency of these categories is predictable. However, among these commonly used words and categories lies 01.16.07.02

Men			Women		
No. in List	Category	Freq.	No. in List	Category	Freq.
12	Knight	174	59	Lady	48
22	Man	104	120	For Women of Rank	32
197	Son	21	157	Woman	25
			216	Queen	20

Table 5: Position and frequency of categories referring to men and women

[All/the whole amount of] in eighth place with 323 occurrences. Looking back at the text, this category is used exclusively for the word "all", which therefore makes the prevalence of this category unexpected. An explanation could be that making sweeping all-encompassing generalisations is a recurring feature in Spenser's style in the first book of *The Faerie Queene*, but this cannot be proven until a further investigation of the collocates and concordance lines is undertaken.

Within the theme of chivalry, there are multiple semantic categories that have relatively high frequencies. As previously mentioned, 03.01.06.01.02.09 [Knight] is very near the top of the list with 174 occurrences, but the question remains as to the context of the word- is it used as a generic term or as a part of the name of the Redcrosse Knight? The proportions of these usages will be discussed in the Discussion section. In the category frequency list, there are attributes favoured in chivalric literature that have a large frequency; 02.02.08-06 [of high/great importance] is found 149 times in the dataset and 03.06.03-05 [fair/equitable] 110 times. Being "great" and "fair" are both desirable descriptors within chivalric literature, so their frequency is to be expected, but some usages of "fair" might be better suited to being a synonym of "beautiful" rather than "equitable". This will become clearer after an examination of concordances.

An interesting comparison that can be made is between different categories referring to men and women. Table 5 illustrates the categories with a frequency higher than twenty that relate to terms for men and women or the roles they occupy. While there are more terms for women, they occupy much lower positions in the list than the categories for men. For example, 01.04.04.01 [Man] is twenty-second on the list, whereas the equivalent female category

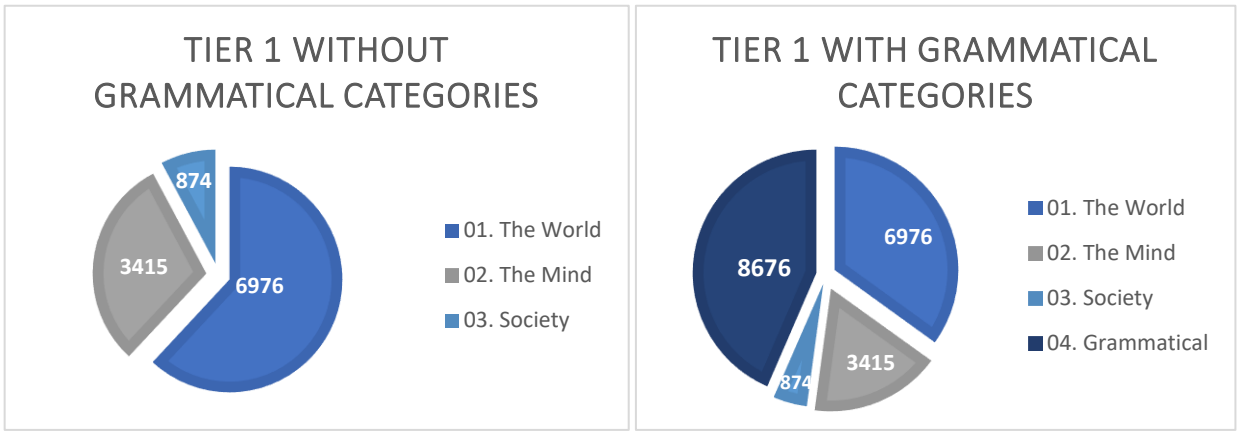


Figure 9: Tier 1 category proportions including and excluding grammatical categories

01.04.04.02 [Woman] is all the way down in 157th place. Even the female category with the highest frequency, 03.01.06.01.02.03-03 [lady], is only in fifty-seventh place. At first glance, these low frequencies suggest that women are simply less important than men in the first book of *The Faerie Queene*, but this conclusion is incorrect. The most important women in the story are Una and Duessa, who are referred to by name and would therefore not appear in this list. The category where they belong, 04.01.01 [Personal Name] – along with the names of the gods and goddesses referenced throughout the story – is found eighty-eight times and is not far below "Man" in twenty-eighth place. Furthermore, as hinted at in Table 5 and confirmed by examining the rest of the category headings, there are more categories that refer to women than men, so the number of references to women are divided among more categories than those of men, bringing the frequencies of individual categories down. This is an important issue in discussing Spenser's style and focuses in *The Faerie Queene* and warrants further investigation.

Beginning with the first tier of categories assigned by the HTST, the importance of grammatical words to the function of English is apparent when examining how the inclusion of grammatical categories affects the proportions of other tier one categories. Excluding grammatical categories, "The World" accounts for approximately two thirds of the instances of categories found in *The Faerie Queene* dataset with 6,976 occurrences. However, with the inclusion of grammatical categories, 6,976 is reduced to only a third of the total number of tier one classifications (see figure 9). The frequent usage of grammatical categories is expected, given that effective discourse requires extensive use of grammatical terms, but it is interesting seeing this distribution illustrated in this way using semantic categories.

Tiers two and three can be discussed together, as it is often the case that subcategories of the most frequent tier two categories are also the most frequent in tier three. As previously mentioned, the most frequent categories in tier two are "Possession", "Relative Properties", "Time", and "Existence and Causation"; these four categories make up 47% of all tier two category usages. In a similar vein to grammatical words, these categories and their tier three subcategories like "Owning", "Existence", and "Relative Time", are integral to effective storytelling in that they help establish relations between objects, settings, events, and characters. Further investigation is needed to establish whether there are any noteworthy conclusions to be drawn beyond these observations.

Following on from these tier two categories is "Life" in fourth place with 761 occurrences in the dataset and the two daughter categories present within "Life" are "The Body" and "Death".[9] Given the amount of combat and vanquishing of enemies within *The Faerie Queene*, the presence of "Death" in this list is to be expected. However, "The Body" is the third most frequently used category in tier three with 513 occurrences, which corroborates the observations made while gathering the data in AntConc. The exact reasons behind his category's prevalence will be explored in the Discussion section.

Another noteworthy observation made in this stage of analysis is the fact that the tier two categories "The Supernatural" and "Morality" have similar frequencies, with 179 and 158 occurrences respectively. This is noteworthy because both categories encompass ideas integral to religion and its treatment in the first book of *The Faerie Queene*: "The Supernatural" includes "Deity" (and also "Supernatural Being", referring to "Fairy/Elf") and "Morality" includes "Duty/Obligation", "Rightness/Justice", and "Moral Evil". In fact, the frequencies of "Deity" (128) and "Rightness/Justice" (113) themselves are comparable. Such similarity in usages suggests that the ideas of religion and morality are intertwined, with duty to God and overtly Christian practices being as important as being a good person. Such interpretations are worth exploring in greater depth with reference to other stages of research.

[9] In the second edition of the Historical Thesaurus, the headings within "Life" have been reorganised so that "Death" is no longer a subcategory of "Life"; they are now both subcategories of the tier two category "Life and Death". Since the semantic tagger uses the first edition of the Historical Thesaurus, "Death" will be treated as a subcategory of "Life" in this research.

5.7.2 Treemap

One of the most striking things about the treemap featuring all the most frequent semantic categories from *The Faerie Queene* dataset is the proportions of the tier-one categories. Within the HT proper, covering the entire history of the English language, "The World" is approximately the same size as "The Mind" and "Society" combined (figure 10). With *The Faerie Queene* treemap, as with the pie charts deriving from the HTST data, "The World" noticeably covers more than half of the graph. This could be due to closed-class categories within "Relative Properties", "Movement", and "Existence and Causation" – which include some of the most common yet irrelevant words (within the context of this research) – having not yet been filtered out. However, there are still a great number of open-class categories present, such as "Sight/Vision", "The Body", "People", and "Matter". The large presence of such categories suggests that Spenser's focus in *The Faerie Queene* is on more tangible and 'real' aspects of existence as opposed to cerebral and philosophical matters and societal constructs and interpersonal relations. Further examination of specific parts of speech will determine if this interpretation stands up to scrutiny and whether this focus on categories within 'The World' is accurate or has in fact been skewed by closed-class parts of speech.

In a similar vein, this treemap shows that "The Mind" is much larger than "Society"; in the HT, these two categories are of a roughly equal size. This is due to pronoun categories like "His/Her" skewing the results and making the "Possession" category seem more important than it is in other parts of speech. Furthermore, the increased size of "the Mind" means that "Society" has been shrunk down to such a size that an in-depth analysis of the categories present is challenging, in addition to diminishing the importance of the data contained therein. Such misrepresentations highlight the significance of filtering the semantic categories in this dataset by part of speech.

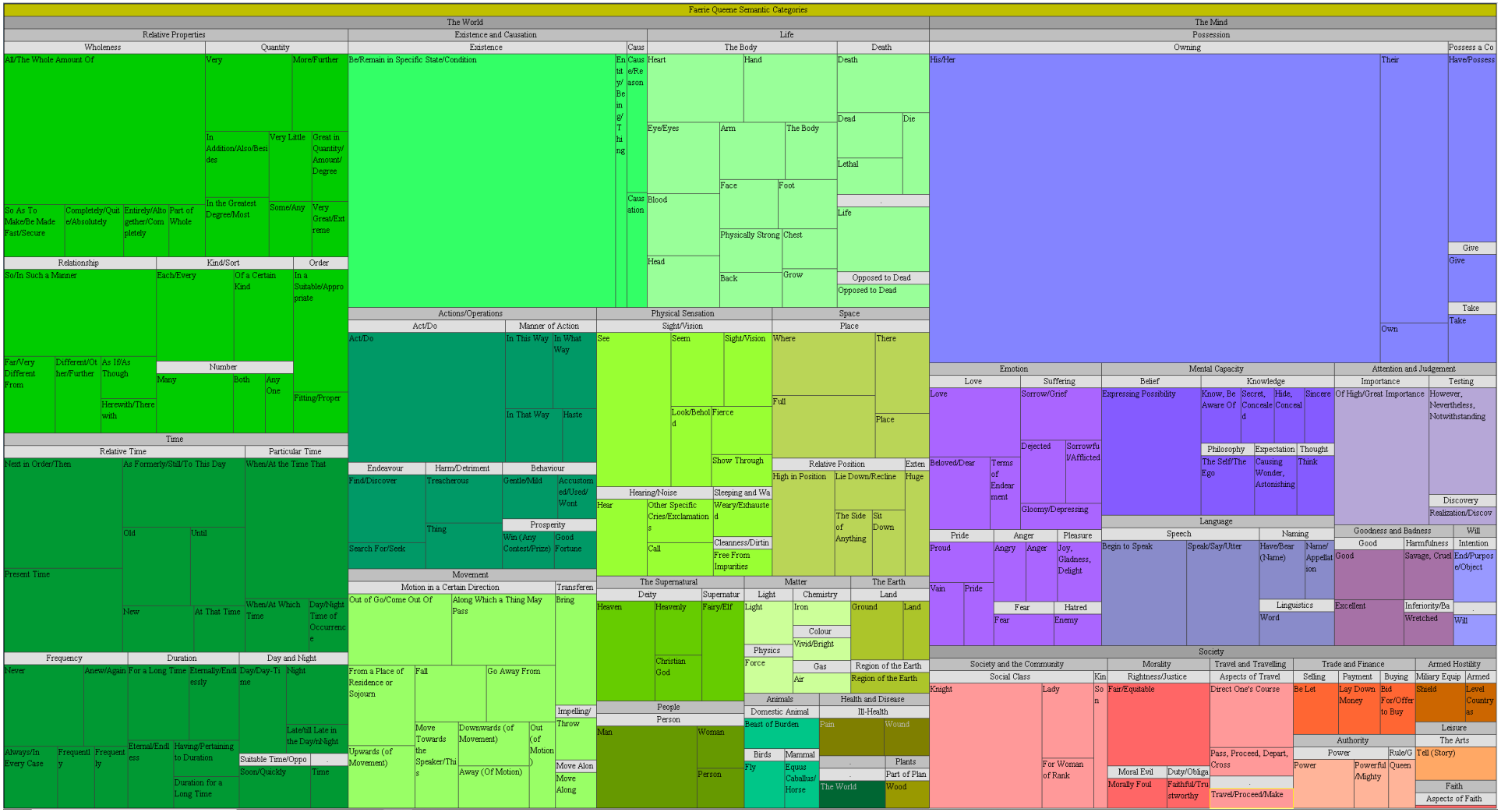


Figure 10: Faerie Queene semantic categories treemap

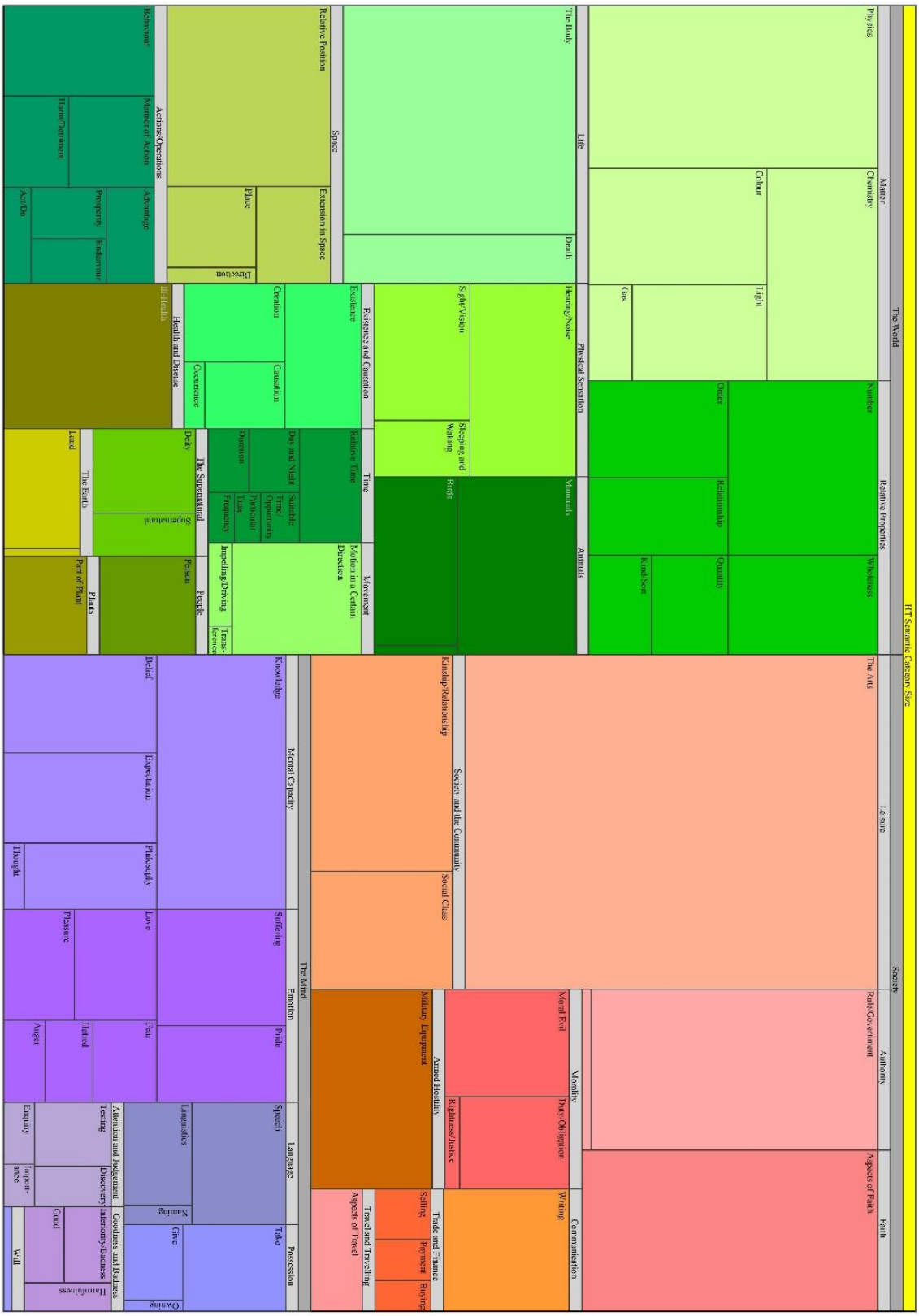


Figure 11: Historical Thesaurus thematic categories treemap

The treemap containing the semantic categories marked in the HT as nouns provides a significantly clearer insight into the semantic tagger results than the larger treemap.[10] In this representation, "The World" is still significantly larger than "The Mind" and "Society" combined, but it is now easier to examine the reasons behind this. The largest tier three category within "The World" is "The Body", with one quarter of the space within "The World" belonging to this category. The significance of words for body parts in *The Faerie Queene* has been hinted at in previous stages of research (see chapter four) but finding confirmation for this emphasis on the body with multiple methodologies is of enormous use in uncovering Spenser's interests.

Another area of significance highlighted in this treemap is the size of the category "Emotion". This category is by far the largest of the four within "The Mind", indicating that the emotions of Spenser's characters receive much of his focus with regards to cognitive matters. The emotions listed can all be treated according to a sixteenth-century viewpoint as either clearly positive, like love and pleasure, or negative, like pride, fear, anger, hatred, and suffering, with none that straddle the line between the two. Such a polarisation of emotions relates to John Duns Scotus' ideas of the senses perceiving objects and experiences as either pleasant or unpleasant, which carries over into emotions (Knuuttila 2004, pp.265–271). Even though Scotus was writing in the thirteenth century, his ideas are still applicable to Spenser's portion of the early modern period (Knuuttila 2012, pp.118–9). This separation between positive and negative emotions suggests that the purpose of emotions in *The Faerie Queene* is not to accurately represent the often-ambiguous emotions of real people, but to serve a symbolic and moralistic purpose within the story. There are also more negative emotions than positive, indicating the need to emphasise evil and immorality in contrast with the potentially assumed default of goodness and morality.

The more detailed outlook afforded by focusing on nouns allows for a clearer view of the "Society" category, which shows "Social Class" to be the largest subcategory of "Society" in this word class. This is to be expected, given the focus on nobility in *The Faerie Queene*, as is the significance of the "Knight" category within "Social Class". However, it is important to note that

[10] The treemaps illustrating the nouns, verbs, and adjectives can be found in Appendix B.

within the story, “knight” does not always refer to the generic title or position but is often included within the character name “Redcrosse Knight”. Distinguishing between these two usages would have been too time-consuming, especially given the frequency of the word in the text, so the usages have not been separated here, resulting in the need to include a caveat when discussing the significance of this category.

As discussed in chapter five, the number of verbs in this dataset is fewer than other open-class word categories and there are even fewer still that have any significant value in the discussion of the semantic categories in *The Faerie Queene*. The verb treemap mostly consists of categories containing common verbs like “Act/Do”, “Go Away From”, and “Be/Remain in a Specific State/Condition”. These verbs are most relevant to the progression of the narrative as they describe the characters’ general movements, rather than specific actions that might have more nuanced descriptions. The treemap illustrating adjectival categories has similar proportions of tier one categories to the one featuring all parts of speech- “The World” is much larger than the sum of the other two, and “Society” is significantly smaller than “The Mind”. However, “Relative Properties” still contains common categories relating to quantity and number like “Very”, “Many” and “Both”. As such, most of “Relative Properties” can be discounted, resulting in “The World” being more similar in size to the HT in relation to the other tier one categories.

Like the nouns treemap, the importance of “Emotion” is clear within the “Mind”, except in the adjective treemap, it does not dwarf the categories of the same tier. The emotions listed were all featured in the noun list and there are still more negative than positive emotions listed. This corroborates the conclusions made above concerning the black-and-white allegorical approach Spenser takes with his characters’ emotions and the emphasis on feelings that evoke immorality in order to supplement his focus on morality.

Morality is a notion detailed in this treemap more than the others, as can be seen in the size of the “Morality” and “Goodness and Badness” categories. It is here that positively connoted categories like “Good” and “Rightness/Justice” can be found in a higher proportion to negative ones like “Moral Evil” and “Harmfulness”. The presence of such a number of categories dealing with these ideas of goodness and morality is juxtaposed with the small size or lack of categories relating to religion; in the adjective treemap, the only category relating to religion is

"Heavenly". Such a contrast suggests that the focus of *The Faerie Queene* is on the notion of morality rather than organised religion, but this idea will be explored in chapter six.

5.8 Limitations and Other Observations

The use of the HTST was one of the more challenging portions of this research due to a number of technical issues. The version of the tagger I had access to was designed for Linux, whereas I only use Windows computers, and attempts to alter the tagger to work on Windows were unsuccessful. Furthermore, there was a lack of technical documentation for troubleshooting the tagger and the readme file did not provide any useful information. To overcome these issues, the txt file of the first book of *The Faerie Queene* that had been normalised was sent to Dr Scott Piao at the University of Lancaster to be input into the version of the semantic tagger there. The resulting tagged file was then returned and used in this research. These logistical limitations are also the primary reason the EEBO reference corpus was not used as a point of comparison for examining how Spenser's semantic distribution compared with Early Modern English as a whole.

Upon examination of the results of the semantic tagger, it became clear that there were numerous errors in the tagging process that needed to be corrected. In addition to the category changes detailed in Table 6, three other categories were removed; one was a duplicate category to one of those changed, and two were mislabelled grammatical categories that did not belong in the HT hierarchy. For example, category 02.06.01-03.02 [denationalize] was a misinterpretation of the pronoun "my", so the category needed to be removed from the data.

With regards to the changes made to the categories listed in Table 6, some errors were the result of the tagger selecting a slightly different sense than the one identifiable from context, such as with "strong" and "son". Others have senses that have nothing to do with the context of the text, like with the misinterpretation of "eye" as an eyelet in clothing or "field" as a magnetic field. Two of the most unusual category misreadings are "true" as 01.12.06-02.01 [measured relative to true north] (instead referring to sincerity and honesty) and the exclamation "O" as 03.09.07.04.01 [Specific written characters]. Out of 207 categories featured more than twenty times, twenty-nine had to be changed or removed, giving the HTST

Lemma	HTST Category	Corrected Category
shield	03.09.04.07.01.01 [Escutcheon/shield]	03.03.16.02.07 [Shield]
living	02.07.01-11 [living/dead/archaic]	01.02-04 [Opposed to dead]
eye	01.08.02.01-18.03.02 [eye]	01.02.03.08.01.04.06 [Eye/eyes]
breast	01.05.11.02.02-02 [breast]	01.02.03.08.03.03-03 [Chest]
son	01.05.09-02.04 [son]	03.01.01.03.02.02 [Son]
wound	01.06.11-06 [wound]	01.03.01.08-05 [Wound]
false	01.06.12-03 [true/false form of]	01.15.17.04-06.02 [treacherous]
gentle	01.01.04.04.01.07-03 [gentle]	01.15.21.04.02.01 [gentle/mild]
fresh	01.01.04.04.04-05 [fresh]	01.09.10.02 [free from impurities]
deadly	01.02.03.09.06.02 [Paleness]	01.02.05.03 [lethal]
steed	01.05.19.06.02.03-06.01 [swift steed]	01.05.19.06.02.01 [Equus caballus/Horse]
field	01.10.07.03.07-04 [magnetic field]	03.03.02.02-02 [Level country as scene of battle]
go	01.11.01.01-01 [end/cease to exist]	01.14.03 [Move along]
way	01.11.01.07-13 [that way/condition]	01.14.05-08.03 [Along which a thing may pass]
away	01.11.02.02-02 [away]	01.14.05.13 [Away (of motion)]
thus	01.11.03.06 [Effect/result]	01.15.20-07 [In this way]
come	01.11.04-03 [happen to]	01.14.05.12-03 [Move towards the speaker/this place]
wide	01.12.05.04-04 [wide open/gaping]	01.16.06.03 [Great in quantity/amount/degree]
oft	01.12.05.25.05-05 [at frequent intervals]	01.13.10 [frequently]
true	01.12.06-02.01 [measured relative to true north]	02.01.12.08.04 [sincere]
strong	01.15.14.02-02 [greatly]	01.02.03.05.01 [physically strong]
eke	01.16.06.05 [Increase in quantity/amount/degree]	01.16.06.05.01-02 [in addition/also/besides]
have	02.01 [Have in the mind]	02.06-05 [Have/possess]
how	02.02.03-02 [as a request]	01.15.20-05 [in what way]
mote	02.02.08.01-02 [of little importance/trivial]	02.01.13.08.09 [Expressing possibility]
O	03.09.07.04.01 [Specific written characters]	01.09.09.02.04.03-02 [Other specific cries/exclamations]

Table 6: Semantic category changes

a success rate of 86%. This is in line with the expected accuracy of the tagger for fiction texts in this approximate time period, as shown in Piao et al.'s analysis of the *Merry Wives of Windsor* (Piao et al. 2017, p.126). Despite having to make changes to some of the categorisations, this is still a reasonably high success rate for the purposes of this research and was more efficient than tagging the text manually.

Using Treemap was a relatively clear way to visualise the data from the semantic tagger, but there are ways in which the data represented could be more comprehensible and effectively displayed. As discussed in his article exploring the uses of Treemap in representing HT data, Alexander (2011, p.6)describes how varying the size of Treemap squares according to entry size is not the most effective way of communicating linguistic data, instead opting to use the software to illustrate diachronic changes across the HT. However, this project is synchronic in nature and the purpose of the treemaps is to provide a complementary visual reference to the HTST data, so the lack of concrete linguistic data in the diagrams is not detrimental to the semantic category analysis.

In the smallest squares – the categories with the lowest frequencies – the headings get cropped out, so you cannot see what the heading is. A similar problem is faced with the categories with long headings- the ends of words are removed to make them fit within the squares. These issues were overcome by manually retyping the headings using image editing software, however having the option to change this within Treemap would have been beneficial.

The specialised nature of the data inputted into Treemap caused a few minor issues with the hierarchy of the HT. There are some instances where words are found under tier one and two category headings, resulting in them being included in the Treemap. However, due to how Treemap parses the input tables, these headings appear to be in tier three with no other categorisations. This was an issue that was unable to be resolved, but it was a relatively minor one as none of the categories affected were particularly sizeable or relevant to the research.

6 DISCUSSION

The purpose of this discussion chapter is to bring together all the data from each stage of analysis into a cohesive examination of key ideas in the first book of *The Faerie Queene* and answer the questions posed at the beginning of this research. First, the focus of the results will lean towards aspects relating to the theme of chivalry such as ideal attributes for men and women in different societal roles. The discussion will then explore how Spenser uses terms relating to the human (and animal) body in the text, followed by an examination of the significance of the emotions Spenser describes. The final theme explored is religion and morality and how Spenser treats these issues, focusing on specific lexical and semantic examples. The conclusions resulting from these analyses will then be used to formulate answers to this project's research questions.

6.1 CHIVALRY, MEN, AND WOMEN

A recurring observation made in each stage of analysis was the strong presence of "knight"; it was the most common noun identified in WMatrix and among the most common semantic categories as demonstrated by the HTST and Treemap. However, the question of which knight was being referred to was asked with each of these findings. The Redcrosse Knight is the protagonist of the first book of *The Faerie Queene*, but he encounters other knights throughout the course of the story, and there is the potential for the word to not refer to any specific knight, but instead be a generic term. Untangling the different usages of "knight" in this portion of *The Faerie Queene* not only gives us an insight into the main character, but the concept of Spenser's representation of knighthood as a whole.

In the first book of *The Faerie Queene*, the word "knight" appears 174 times, 21 of which are in collocation with "Redcrosse", indicating that the Redcrosse Knight is being specified. Through investigating the collocates and concordances of the remaining 153 occurrences, it becomes clear that sometimes knights are distinguished by particular attributes or descriptions. Examples of this include "recreant knight", "redoubted knight", "errant knight", and "doughty knight". The majority of these examples are more generic references to knights, such as "Strange thing it is an errant knight to see/Here in this place" (Spenser 1596, ll.4719–

20) or as an alternative name for a named knight; “courteous knight” is used in reference to the Redcrosse Knight. By defining knights by a single adjective at a time, Spenser is reducing the knight to a single attribute given as if it were his name. These epithets are a defining feature of Spenser’s style (Webster 1976, pp.80–81) and, in this instance, could be treated as simplifying knighthood to a handful of relevant (if not necessarily desirable) characteristics.

The most significant collocates listed for “knight” are not the same as those for “Redcrosse Knight”, though. As previously discussed, among the top collocates for “knight” where “redcrosse” appears within five words at either side are “gentle”, “love”, and “good”. At first glance, this suggests that the Redcrosse Knight is a different sort of knight to those mentioned previously, full of compassion and chivalric kindness. However, upon further examination of the concordances, this interpretation is not entirely accurate.

1	him so many wrathful wreaks: / For never	**gentle**	knight, as he of late, / So tossed
2	it is, that erst did throw / This	**gentle**	knight into so great distress, / That death
3	ran through every joint, / For ruth of	**gentle**	knight so foul forlore: / Which shaking off,
4	faint and feeble grow. / Hereof this	**gentle**	knight unwitting was, / And lying down upon
5	Doth to his home entreat. / A	**gentle**	Knight was pricking on the plain,
6	Ah Lady dear, quoth then the	**gentle**	knight, / Well may I ween, your grief
7	yet again, and yet again bespake / The	**gentle**	knight; who nought to him replied
8	that Trevisan had told, / When as the	**gentle**	Redcrosse knight did view, / With fiery zeal
9	ground. / Thine, O then, said the	**gentle**	Redcrosse knight, / Next to that Ladies love,

Concordance 1: All concordances of 'gentle' in proximity to 'knight'

With “gentle”, as illustrated above, there are multiple references to the Redcrosse Knight, both explicit and covert. Lines 8 and 9 are two examples where the Redcrosse Knight is directly described as “gentle”, with the first including the phrase “fiery zeal” in its character description. Other examples demonstrate a slightly more obscure reference to the Redcrosse Knight, like line 5 with the well-known sentence “A gentle knight was pricking on the plain” and line 4 where the Redcrosse Knight drinks magic water that saps his strength and weakens him. Upon examining the passages the concordances are drawn from, it becomes clear that “gentle knight” functions as an alternative name for the Redcrosse Knight, which can be seen as reflecting the text’s allegory (Hamilton 2019). It is not the case, however, that a specific term is used in specific contexts; the examples in Concordance 1 come from romantic situations (lines 6 and 9), as well as times of distress (lines 2, 4, and 7) and misfortune (line 1). Therefore,

it seems that gentleness is an integral characteristic to the Redcrosse Knight and not a facet that comes and goes depending on the situation he is faced with. The ideas of *gentilesse*, behaving in an honourable manner, and the nobility to which the Redcrosse Knight belongs are all firmly rooted in medieval culture and chivalric literature, and gentility is one of the ultimate virtues a knight can aspire to (Keen 2005, p.160). Not only is Spenser harking back to this tradition by using this descriptor, but he is also indicating that the Redcrosse Knight is an ideal example of knighthood for having reached the peak of chivalric behaviour.

1	knew well all was true. But the	**good**	knight / Full of sad fear and ghastly
2	the person to put on / Of that	**good**	knight, his late beguiled guest: / In mighty
3	(quoth she) I hither came, / And this	**good**	knight his way with me addressed, / Led
4	foully shamed, and doest vaunt / That	**good**	knight of the Redcrosse to have slain:
5	that with thy cruel dart / At that	**good**	knight so cunningly did rove, / That glorious
6	shunned the unlucky ground. / But this	**good**	knight soon as he them can spy,
7	as one of the train: / But that	**good**	knight would not so nigh repair, / Him

Concordance 2: All concordances of 'good' in proximity to 'knight'

With regards to the collocates of “good”, as illustrated in Concordance 2, there is one clear reference to the Redcrosse Knight in line 4. In a similar vein to “gentle”, the other quotations seem to reference the Redcrosse Knight within the larger context of the story, but these references are less clear cut than with “gentle”, only becoming clear when looking at the placement of the phrases within the wider text and examining the surrounding stanzas. In fact, upon looking at the wider text, line 7 most likely refers to the “elfin knight”. Such vagueness about who exactly is a “good knight” suggests that the attribute refers to knights more generally, as opposed to being another name for specific knights. The concordances of “love” in proximity to “knight” (Concordance 3) indicate that the use of “love” is different to that of “gentle” and “good”. Line 1 technically refers to the love of God but reads more as an exclamation. Line 3 refers to the Redcrosse Knight being loved rather than him being the one feeling love, which makes it irrelevant as a possible knightly attribute. Line 2’s entire quote reads:

> “For unto knight there is no greater shame,
> Then lightness and inconstancy in love;
> That doth this Redcrosse knight's ensample plainly prove.” (Spenser 1596, ll.1519–21)

1	forth to shake. / For Gods dear **love**, Sir knight, do me not stay; / For
2	Then lightness and inconstancy in **love**; / That doth this Redcrosse knight's ensample
3	Lady loved dearer day, / Then she did **love** the knight of the Redcrosse; / For whose

Concordance 3: All concordances of "love" in proximity to "knight"

This quote is part of a speech given by Duessa aimed at provoking the Redcrosse Knight for his confused feelings towards Una and Duessa in her many disguises. While this might suggest that the Redcrosse Knight's flaw is his inability to see beyond the tricks Duessa plays on him and control his feelings of love towards any woman that crosses his path, the fact that this speech is given by one of the main antagonists in the story could devalue this interpretation. Duessa is a known liar, so she could be lying about love and faithfulness being an integral part of knighthood. However, she could also be manipulating him by taunting him with a characteristic the knight does not have. Whichever interpretation is correct, the result of this is that the evidence for love being a desirable chivalric trait is very minimal, despite the frequency of the concept, with the focus instead being on goodness and gentleness.

Among the initial observations for the data from the HTST was the differences between categories relating to men and women in terms of frequency. At first glance, it seemed that while there were fewer categories relating to men, they were used more frequently than the categories relating to women, of which there were more in number.[11] These observations were worthy of closer examination, and the most detailed way to examine this was to look at the frequencies of the individual words within these semantic categories.

Table 7 gives a list of all the words for men and women in the first book of *The Faerie Queene* that have a frequency higher than ten. Previously, the minimum relevant frequency has been twenty, but that yielded too few words to make any noteworthy observations. The log-likelihood value derives from the WMatrix data, where the frequencies of words in *The Faerie Queene* dataset were compared against those from a reference corpus consisting of similar texts from the same period. A log-likelihood value higher than 6.63 is considered significant. The column labelled "+/-" indicates whether a word has a higher relative frequency in *The Faerie Queene* or EEBO dataset; a plus sign indicates that the word is found more in the first

[11] See Table 5 in chapter five for a full list of these categories.

Men	Freq.	LL	+/-	Women	Freq.	LL	+/-
knight	169	338.47	+	*lady*	55	66.42	+
man	66	2.15	-	*Una*	43	163.31	+
men	32	29.87	-	*dame*	36	33.51	+
lord	29	1.88	-	*Duessa*	31	127.38	+
sire	21	23.72	+	*queen*	27	1.52	-
son	19	1.89	+	*daughter*	21	5.26	+
king	18	29.38	-	*ladies*	21	8.28	+
sir	17	5.33	+	*woman*	14	9	+
knights	17	28.33	+	*maid*	13	8.79	+
prince	13	1.72	-	*mother*	10	1.42	+
squire	11	8.47	+	*witch*	10	26.39	+
Sansfoy	10	45.73	+				

Table 7: Frequencies and log-likelihood values for words for men and women

book of *The Faerie Queene,* whereas a minus sign means it is more common in the EEBO reference corpus.

As previously discussed, the HTST data indicated that there was greater variety in categories used for women, but the categories for men had higher frequencies. However, the data from WMatrix refutes this: not only are there more words for men, but their frequencies tend to be higher than words for women and have a higher overall frequency; the total frequency for masculine words is 281 compared with 422 for feminine. However, things get more interesting when the EEBO reference corpus is taken into account. Words for women are overused in *The Faerie Queene* dataset compared with the reference texts in a way that the words for men are not; the male words that are marked as overused – "knight(s)", "sire", "squire", "Sansfoy", and "son" – can mostly be explained by the subject matter of *The Faerie Queene*, with its abundance of knights and other such noble roles. This is not the case with the words for women, where terms for all sorts of women are used more in *The Faerie Queene* dataset, with the one exception of "queen". Given the number of texts in the EEBO reference corpus that mention Queen Elizabeth I, this is unsurprising. Even when comparing the log-likelihood values, there are more statistically significant words relating to women than men. Despite the fact that men are mentioned more than women in the first book of *The Faerie Queene*, women still have a larger presence in the story than in other contemporary texts. This could suggest that Spenser is more progressive than his contemporaries in terms of including more women in his work. It

could also suggest that women are necessary to the genre of chivalric literature; Emelye is the source of the dispute between Palamon and Arcite in Chaucer's *Knight's Tale,* and much of the plot of Chrétien de Troyes' *Le Chevalier au Lion* is based around Yvain trying to win back Laudine. However, the only reason that Laudine rejects Yvain in the first place is that he forgets to return to her after a year because he is too busy questing, which suggests that Laudine has lots of agency in the story. With *The Faerie Queene,* however, Anderson argues that women play a more integral role to the story than in other early modern chivalric tales and have a stronger sense of agency (2018, p.38), which could make *The Faerie Queene* more like earlier examples of chivalric literature than early modern. Such an interpretation indicates that Spenser's treatment of women in this section of *The Faerie Queene* is an aspect of his style that marks him out as different to his contemporaries.

6.2 The Body

The importance of words relating to parts of the body is a semantic category that has been flagged up at multiple stages of analysis as significant; an unexpectedly high frequency of body part terms was noticed during the WMatrix and AntConc stages of data collection, which was corroborated by the frequent usage of the Historical Thesaurus category "The Body" as identified by the semantic tagger. To gain a deeper understanding of why these words are so common and how Spenser uses them, a comparison with the EEBO reference corpus will be made, followed by an in-depth examination of the collocates of each body part word.

Lexeme	Freq.	+/-	LL	Lexeme	Freq.	+/-	LL
back	19	+	35.43	*face*	33	+	10.58
arms	34	+	29.71	*foot*	12	+	8.79
blood	57	+	22.55	*eyes*	46	+	5.19
hand	48	+	16.35	*heart*	69	+	5.07
breast	25	+	14.12	*head*	33	+	2.17
feet	20	+	12.95	*hands*	20	+	1.43
eye	31	+	11.83	*body*	22	+	1.43
				flesh	15	-	0

Table 8: Body part word frequencies and log-likelihood values

For the frequency comparison with the EEBO reference corpus, the words from each semantic category within the tier three category "The Body" were identified and recorded. These words were subsequently identified in the data from the comparison between *The Faerie Queene* and EEBO corpora in the same manner as the words relating to men and women. Table 8 illustrates the frequency of the body part words and whether these words are over- or underused within *The Faerie Queene* corpus, in addition to the log-likelihood values for each word. A special note must be made about "back"- since WMatrix does not distinguish between senses or part of speech in its word keyness analysis, "back" here refers to not only the body part but also adverbial and other noun senses. Examining the semantic category data shows that such broadness of scope must be taken into account when examining the significance of this word but is not an issue for the other words discussed.

The comparison with the EEBO reference corpus demonstrates that almost all the body part terms are found more in *The Faerie Queene* dataset than the EEBO texts used in the reference corpus with the exception of "flesh". It is unclear why "flesh" is so prevalent in the reference corpus without a close examination of the texts included, but it is possibly due to the inclusion of religious texts, since "flesh" is commonly used in Elizabethan religious discourse to differentiate the fallible corporeal body from heavenly concerns (Narveson 1999, p.322; Borris 2000, p.21). As for the overuse of the other terms, this confirms the observation that the first book of *The Faerie Queene* includes an unusually high number of body part terms, but the question as to why still remains.

Upon closer examination of the collocates of each of the body parts listed in Table 8, it becomes clear that there are three main functions of body part words: the first is to refer to a literal part of the human body, the second a representation of a concept, and the third as a means of developing characterisation. Some words embody different functions in different contexts, like "heart" and "blood", so each function will be discussed separately.

A handful of the words analysed refer simply to the body part they are named for. One of the most frequent uses of "blood" is to describe the appearance or characteristics of blood, usually emanating from a wound (see lines 1, 8, 14-16, 18, and 20). Sometimes Spenser uses "blood" in hyperbolic phrases within violent scenes, such as "large floods of blood" (line 15) and "a sea

of blood" (line 16). The purpose of this is to create evocative images in the mind of the reader, and it is notable that a part of the human body is involved in such aspects of style.

1	without remorse; / A stream of coal black	**blood**	forth gushed from her corpse. / Her
2	poisoned garment did enchant / With Centaurs	**blood**,	and bloody verses charmed, / As did this
3	cold his courage began assail, / And cheerful	**blood**	in faintness chill did melt, / Which like
4	cunning imagery, / On which true Christians	**blood**	was often spilt, / And holy Martyrs often
5	luckless fray. / His cruel wounds with cruddy	**blood**	congealed, / They bind up so wisely, as
6	hands from guilt of bloody field: / For	**blood**	can nought but sin, and wars but
7	he had no government, / Ne cared for	**blood**	in his avengement: / But when the furious
8	rift there came / Small drops of gory	**blood**,	that trickled down the same. / Therewith
9	rare, / And rueful plaints, me bidding guiltless	**blood**	to spare? / Then groaning deep, Nor
10	Whose manly hands imbrewed in guilty	**blood**	/ Had never been, never by his
11	whose defence he was to shed his	**blood**.	/ At last dull weariness of former fight
12	With thine own blood to price his	**blood**,	here shed in sight. / What frantic
13	That happy land, and all with innocent	**blood**	/ Defiled those sacred waves, it rightly hot
14	bight / In tender flesh, that streams of	**blood**	down flow, / With which the arms, that
15	pity any living eye. / Large floods of	**blood**	down their sides did rail; / But floods
16	shape miss-shaped more: A sea of	**blood**	gushed from the gaping wound, / That her
17	long fight, / And faint through loss of	**blood**,	moved not at all, / But lay as
18	did prick and nip, / That drops of	**blood**	thence like a well did play; / And
19	suddenly, / Hunting full greedy after salvage	**blood**;	/ Soon as the royal virgin he did
20	teeth enranged were, / In which yet trickling	**blood**	and gobbets raw / Of late devoured bodies
21	a leaf of Aspen green, / And troubled	**blood**	through his pale face was seen / To
22	wallowed in his own yet luke-warm	**blood**,	/ That from his wound yet welled fresh

Concordance 4: Selected concordances of "blood"

1	Else never could the force of fleshly	**arm**,	/ No molten metal in his blood imbrue:
2	heart, is hard to tell: / Upon her	**arm**	a silver anchor lay, / Whereon she leaned
3	him failed long ago, / And on his	**arm**	a bunch of keys he bore, / The
4	staring eye, / He down let fall his	**arm**,	and soft withdrew / His weapon huge, that
5	Then to him stepping, from his	**arm**	did reach / Those keys, and made himself
6	earthly ground: / And who most trusts in	**arm**	of fleshly might, / And boasts, in beauties
7	burning bright / He smote off his left	**arm**,	which like a block / Did fall to
8	beheld, his mighty shield / Upon his manly	**arm**	he soon addressed, / And at him fiercely
9	most goodly virgins came in place, Ylinked	**arm**	in arm in lovely wise, / With countenance

Concordance 5: Concordances of "arm"

Other terms that are used literally in reference to parts of the body are “arm” and “arms”. “Arm” is most often used in phrases that describe something on an arm or a function they perform, such as “upon her arm a silver anchor lay” (line 2 of concordance 5), “on his arm a bunch of keys he bore” (like 3), and “ylinked arm in arm” (line 9). The only unusual pattern seems to be that describing an arm as “fleshly” is only used in contexts where there is – or potentially will be – a battle, but since this is only seen twice no real conclusions can be drawn from this (lines 1 and 6). Otherwise, it is clear that Spenser uses “arm” to mean just that, an arm.

1	band / Of tall young men, all able	arms	to sound, / But now they laurel branches
2	bears a charmed shield, / And eke enchanted	arms	, that none can pierce, / Ne none can
3	rope he wears, / That with his glistering	arms	does ill agree; / But he of rope
4	as he from far descried / Those glistering	arms	, that heaven with light did fill,
5	her nigh weary wane, and in her	arms	/ To Æsculapius brought the wounded knight
6	endure so cruel case, / But thought his	arms	to leave, and helmet to unlace.
7	But had he been, where erst his	arms	were lent, / the Enchanter vain his error
8	his well deserved name: / He had in	arms	abroad won mickle fame, / And filled far
9	Whereas an errant knight in	arms	yclad, / And heathenish shield, wherein
10	on the plain, / Y clad in mighty	arms	and silver shield, / Wherein old dints of
11	A knight her met in mighty	arms	embossed, / Yet knight was not for all
12	that ever loosely led, / Without regard of	arms	and dreaded fight: / Those two he took,
13	spear in warlike hand, / And feats of	arms	did wisely understand. / Eftsoons he pierced
14	bind / to observe the sacred laws of	arms	, that are assigned. / At last forth
15	leave so dearly bought? / What need of	arms	, where peace doth always remain,
16	goodly lodge, and began despoil / Of puissant	arms	, and laid in easy bed; / His name
17	goodly trees, that far did spread / Their	arms	abroad, with gray moss overcast
18	For this young Prince, when first to	arms	he fell; / But when he died, the

Concordance 6: Selected concordances of "arms"

In this text, “arms” has much more varied uses than its singular counterpart. Of the thirty-four occurrences of “arms”, only three refer to the parts of the body; line 17 refers to the arms of trees but is not included in this figure. Of those three, two refer to carrying things or people, thus reinforcing the practical uses of some of the body parts referenced by Spenser. However, there are other senses of “arms” represented in the above concordances. “He had in arms abroad won mickle fame” (line 8) most likely refers to the sense “armed combat as a

1	as brightest sky, / And burnt his beastly	**heart**	to efforce her chastity. / So when
2	not resist, nor succour call, / His bleeding	**heart**	is in the venger's hand, / Who
3	filthiness, / And in his hand a burning	**heart**	he bare, / Full of vain follies, and
4	were often heard to groan, / That hardest	**heart**	would bleed, to hear their piteous moan.
5	death, when she had marked long, / Her	**heart**	began melt in great compassion
6	great mothers face: / Yet pity in her	**heart**	was never proved / Till then: for evermore
7	Did much immove his stout heroic	**heart**	, / And said, Dear dame, your sudden overthrow
8	That glorious fire it kindled in his	**heart**	, / Lay now thy deadly Heaven bow apart,
9	iron brand, / And launched his Lordly	**heart**	; with death oppressed / He roared aloud
10	and lustful play, / That nigh his manly	**heart**	did melt away, / Bathed in wanton bliss
11	Ne let vain words bewitch thy manly	**heart**	, / No devilish thoughts dismay thy constant
12	begins for to creep, / And in my	**heart**	his iron arrow steep, / Soon as I
13	so dear as life is to my	**heart**	, / I deem your love, and hold me
14	this lesson dear / Deep written in my	**heart**	with iron pen, / That bliss may not
15	delude, or true it were, / Was never	**heart**	so ravished with delight, / No living man
16	little comfort shows: / At last recovering	**heart**	, he does begin / To rub her temples,
17	have great cause of plaint, / That stoutest	**heart**	, I ween, could cause to quake.
18	did contend, / And almost rent her tender	**heart**	in tway; / And love fresh coals unto
19	dead. / That cruel word her tender	**heart**	so thrilled, / That sudden cold did run
20	kill, / And raise again to life the	**heart**	, that she did thrill. / And when
21	strife, / Pain, hunger, cold, that makes the	**heart**	to quake; / And ever fickle fortune rages
22	hour, / But in eternal woes my weaker	**heart**	/ Have wasted, loving him with all my
23	and but the stump him left. /	**Heart**	cannot think, what outrage, and what cries

Concordance 7: Selected concordances of "heart"

professional activity" and "y clad in mighty arms and silver shield" (line 10) to the sense "defensive coverings for the body worn when fighting; armour" (arms, n. 2021). All of these military senses have a completely different etymology to the body part, deriving from Latin by way of Anglo-Norman and Old French rather than Germanic (arm, n.1 2021). Such variation in etymology and senses means that any relationship drawn between the military and body part senses is an error on the part of the reader, but these senses do demonstrate a literal usage of words rather than metaphorical.

The second function of body part terms is as a metaphorical representation of a concept. There is another use of "blood" that fits this type and can be seen in the phrases "both for her noble blood, and for her tender youth" (Spenser 1596, l.533) and "did though behight me borne of English blood?" (Spenser 1932, l.5264). The relationship between "blood" and lineage or ancestry is a common connection, present in cognates across Indo-European languages (blood,

1	sharper edge did feel, / Or his baptised	**hands**	now greater grew; / Or other secret virtue
2	nowhere can rest, / But with his clownish	**hands**	their tender wings / He brushes oft, and
3	devise, / How to advance with favourable	**hands**,	/ As this good Prince redeemed the
4	from her to turn away; / With folded	**hands**	and knees full lowly bent / All night
5	was heard, / Crying, O spare with guilty	**hands**	to tear / My tender sides in this
6	and tempt her feigned truth. / Wringing her	**hands**	in women's piteous wise, / Though can
7	his foot was pight; / Then to his	**hands**	that writ he did betake, / Which he
8	her weary feet, / And licked her lily	**hands**	with fawning tongue, / As he her wronged
9	a fresh unproved knight, / Whose manly	**hands**	imbrewed in guilty blood / Had never been,
10	withhold, / (Entire affection hates nicer	**hands**)	/ But that with constant zeal, and courage
11	catching hold, did her dismay / With quaking	**hands**,	and other signs of fear: / Who full
12	mind, / And eke as pledges firm, right	**hands**	together joined. / Prince Arthur gave
13	freight, / Bade on that Messenger rude	**hands**	to reach. / Eftsoons the Guard, which on
14	fell from high honours stair, / Into the	**hands**	of his accursed fone, / And cruelly was
15	glee, / Shouting, and clapping all their	**hands**	on high, / That all the air it
16	of earthly conquest shone, / And wash thy	**hands**	from guilt of bloody field: / For blood
17	ever to abide. / His own two	**hands**	the holy knots did knit, / That none
18	of time, and everlasting fame, / That warlike	**hands**	ennoblest with immortal name; / O gently

Concordance 8: Selected concordances of "hands"

n. (and int.) 2021). Nonetheless, Spenser is still using a part of the human body to represent a non-corporeal concept, and “blood” is not the only word he uses in this way. The first book of *The Faerie Queene* does have some references to the heart as an organ, like in line 12 of Concordance 7, but a large portion of the uses treat the heart as the emotional centre of the body. For example, in the line “that hardest heart would bleed, to hear their piteous moan” (line 4), the heart represents compassion, which is explicitly corroborated in line 5. Such conceptual mappings are common in English literature, but their presence throughout *The Faerie Queene* marks out Spenser’s use of corporeal terms to represent more abstract ideas as a feature of his style.

Finally, the third use of words for parts of the body is to further character development. Across the board, Spenser uses a variety of adjectives before body part terms, whose connotations add to the audience’s understanding of the body part’s owner. In Concordance 7, the use of “beastly” before “heart” (Concordance 7, line 1) furthers the characterisation of Sansloy as an evil character. However, this use of body part words is best exemplified with “hands” and “breast”. With “baptised hands” (Concordance 8, line 1), Spenser is reaffirming the knight’s miraculous rebirth after being severely wounded in battle, as well as tying in the knight’s purity

1	in the fire, / Had riven many a	**breast**	with pikehead square; / A goodly person
2	And silver shield: upon his coward	**breast**	/ A bloody cross, and on his craven
3	Can high advance his broad discoloured	**breast**,	/ Above his wonted pitch, with countenance
4	O gently come into my feeble	**breast**,	/ Come gently, but not with that mighty
5	fort can be so strong, / No fleshly	**breast**	can armed be so sound, / But will
6	Did search, sore grieved in her gentle	**breast**,	/ He so ungently left her, whom she
7	fierce encounters fit. / But on his	**breast**	a bloody Cross he bore, / The dear
8	But soon as breath out of his	**breast**	did pass, / That huge great body,
9	Some lingering life within his hollow	**breast**,	/ Or in his womb might lurk some
10	despite; / For never felt his imperceptible	**breast**	/ So wondrous force, from hand of living
11	out of his body grew, / An iron	**breast**,	and back of scaly bras, / And all
12	joy of chivalry / First kindled in my	**breast**,	it was my lot / To love this
13	with mickle smart: / For since my	**breast**	was launched with lovely dart / Of dear
14	life nigh crushed out of his panting	**breast**:	/ No power he had to stir, nor
15	night / whilom doth rankle in my riven	**breast**,	/ With forced fury following his behest,
16	sorrowful regret, / And lively breath her sad	**breast**	did forsake, / Yet might her piteous heart
17	did pass, / Forelifting up aloft his speckled	**breast**,	/ And often bounding on the bruised grass,
18	roared aloud, while life forsook his stubborn	**breast**.	/ Who now is left to keep
19	and evermore does steep / Her tender	**breast**	in bitter tears all night, / All night
20	And seizing cruel claws on trembling	**breast,**	/ Under his Lordly foot him proudly hath

Concordance 9: Selected concordances of "breast"

of religion. Good female characters have descriptions such as “folded hands” (line 4), “lily hands” (line 8), and “quaking hands” (line 11), emphasising the importance of piety, gentleness, empathy and – to an extent – fragility. By contrast, men’s hands are described as “warlike” (line 18) and “manly” (line 9) for the knights and “clownish” (line 2) for a shepherd. These collocates supplement existing character descriptions – like gentle and frail women, brave but violent knights, and silly forgettable peasants – in a similar way to the epithets with character names, thus reinforcing these patterns of descriptions as an important part of Spenser’s style.

Spenser’s treatment of “breast” is similar to that of “hands”: while there are references to literal chests (Concordance 9, lines 1, 7, 8, and 13), there are still numerous examples where the adjectives used with “breast” help to reinforce character traits. “Coward breast” (line 2) reiterates to the reader that Sansfoy is a ‘bad’ character, or at least not one to root for. “Feeble breast” (line 4) is part of a speech spoken by Una where she tries to comfort the knight and get him to calm down, so the use of “feeble” is to underline her feminine qualities; “tender breast” (line 19) serves the same function. Similarly, “sad breast” (line 16) is an excuse for Spenser to

hammer home Una's misery in a way he does not for the male characters. To emphasise Archimago's strength but in a more negative way than the Redcrosse Knight, his chest is described as "stubborn" (line 18). All of these descriptions add further elements to Spenser's characterisation, with some more jarring than others, however this aspect of his style is integral to the development of the vivid characters he creates in *The Faerie Queene*.

6.3 EMOTION

The significance of emotional vocabulary in the first book of *The Faerie Queene* was established during the collocation and semantic category analyses in chapter five. Additional observations about the range of emotions described have also been made, facilitated by the Treemap visualisation of the HTST data. However, questions remain about what the most important emotions are in the first book of *The Faerie Queene*, and how Spenser treats these emotions within the text.

According to the treemaps generated from the HTST data, Love and Suffering are tied for first place in the list of most often-used emotions by Spenser in *The Faerie Queene*. This is in spite of "Love" having three subcategories represented (Love, Beloved/Dear, and Terms of Endearment) and "Suffering" four (Sorrow/Grief, Sorrowful/Afflicted with Grief, Dejected, and Gloomy/Depressing). However, a deeper examination of these subcategories sheds a new light on this interpretation of the data. Table 9 illustrates each word attested in the first book of *The Faerie Queene* from the aforementioned subcategories, and it is clear that Suffering has by far and away the most lexical variety of all the emotions; the number of words used to describe aspects of suffering (18, if both instances of "sad" are counted as one) is double that to describe aspects of love (9). The importance of chivalry to *The Faerie Queene* and the relationships between the Redcrosse Knight and the female characters explains the high frequency of love, but in order to explain the prevalence of suffering, one must look to Protestantism. Suffering is integral to the Protestant psyche, as you must have experienced suffering in order to experience redemption (Ryrie 2013, pp.35–6). Spenser's own Protestantism, as well as his exposure to other branches through other members of his discourse community, means that it is to be expected that more everyday aspects of the Protestant faith beyond overt references are featured in *The Faerie Queene.*

T3 Emotion	Category Heading	Words
Pleasure	Joy/Gladness/Delight	joy, joyous, delights, joyance
Suffering	Sorrow/Grief	grief, ruth, rue, sorrow, sorrows, dolours, teen
	Sorrowful/Afflicted with Grief	woeful, sore, sad, sorrowful, baleful, distressed
	Dejected	sad, dreary, glooming
	Gloomy/Depressing	mournful, darksome, doleful
Anger	Anger	wrath, ire, anger, choler, heaviness
	Angry	angry, wrathful, enraged
Love	Love	love, loves, loved, loving
	Beloved/Dear	dear, dearest, loved
	Terms of Endearment	dear, darling
Hatred	Enemy	enemy, enemies, foe, foeman, foemen, fiend
Pride	Pride	pride
	Proud	proud, prouder
	Vain	vain
Fear	Fear	fear, fears

Table 9: HT emotion categories and words

Now that Spenser's most frequently used emotions have been established, it is now worth exploring exactly how he treats these words in this part of *The Faerie Queene*. Of all the words from 02.04.13 [Love] found in the text, fifty-two are nouns and twenty-three are verbs. Such a distribution could mean that love is more an entity or object than an action that someone performs. To confirm this hypothesis, the concordances of love must be taken into account. Of the fifty-six concordances listed in AntConc, sixteen of the occurrences of "love" are found preceded by a possessive pronoun or a genitive construction like "ladies' love", which corroborates the idea that love is something a person can own or possess. However, there are multiple examples where love is used as the subject of a sentence, such as "which love has launched with his deadly darts" (Spenser 1596, l.4365) and "love, lay down thy bow" (Spenser 1932, l.4161). He is most likely referring to Eros, described in Plato's *Symposium* as the Greek god of love and desire (Moulton 2014, p.29), but even if these phrases are read literally, love is being treated as an external force that is acting on the characters without their input. All

Category	Word	PoS	Freq.	Category	Word	PoS	Freq.
Sorrow/Grief	grief(s)	N	28	*Dejected*	sad	Aj	23
	ruth	N	5		dreary	Aj	8
	rue	N	1		glooming	Aj	2
	sorrow(s)	N	19	*Gloomy/ Depressing*	mournful	Aj	7
	dolours	N	3		darksome	Aj	12
	teen	N	2		doleful	Aj	8
Sorrowful/ Afflicted with Grief	woeful	Aj	21				
	sore	Aj	30				
	sad	Aj	35				
	sorrowful	Aj	3				
	baleful	Aj	9				
	distressed	Aj	2				

Table 10: Frequencies of words in subcategories of "Suffering"

these usages of love in the first book of *The Faerie Queene* demonstrate that love has multiple roles to play within the text and Spenser is making full use of this variety, despite not varying his vocabulary.

The way that Spenser treats suffering is different to the way he treats love. As already discussed, there is a greater variety of words relating to suffering than love, but not every word is equal. Table 10 lists the frequencies of each word found within the subcategories of 02.04.11 [Suffering], and what is clear is that some words are used considerably more than others. Of all the nouns within “Sorrow/Grief”, “grief(s)” and “sorrow(s)” are used 48% and 33% of the time respectively. “Woeful”, “sore”, and “sad” are the most used within “Sorrowful/Afflicted with Grief, making up three-quarters of all uses, and within “Dejected”, “sad” is used 70% of the time. The effect is not as dramatic within “Gloomy/Depressing” due to the overall frequencies being lower, but “darksome” still comprises 44% of occurrences of this category.

As for how Spenser uses these words, there are some noteworthy observations. “Sorrow” is often preceded by adjectives like “fervent”, “great”, “secret”, and “thrilling”, which are unexpected words to be found in combination with “sorrow” to modern readers. These additional descriptors add nuance and prevent the many readings of “sorrow” from being

interpreted the same way by the reader. Furthermore, “woeful” is mostly used in reference to people; “Una”, “daughter”, “thrall”, “dwarf”, and “lady” are all described at one point as being “woeful”. “Sad” is treated in a similar way, being used to describe both heroic and evil characters in addition to objects. What is most interesting about these usages is that they are egalitarian- both good and bad characters are described as “woeful”, indicating that this aspect of suffering has nothing to do with your inherent morality. It is neither a positive nor negative attribute, but merely an aspect of humanity.

In summary, the vocabulary used by Spenser to describe suffering and love is varied both in terms of lexis and usages. Despite the fact that there are few distinct words used to describe love, Spenser’s differentiation between love enacted by the characters themselves and love that is thrust upon them by an external force adds nuance to the romantic subplots. With suffering, there are words that are used more than others, but once again, Spenser’s use of descriptors and subjects adds variety and distinction between repeated usages of words.

6.4 Religion and Morality

As noted in previous chapters, there is a strong focus on religion within the first book of *The Faerie Queene*, no doubt because of Spenser’s own views on Protestantism and his intended audience. However, the semantic category analysis has flagged that words and categories relating to morality are as frequent as overtly Christian terms. This section will explore Spenser’s use of these categories and exactly how he treats Protestant ideals in this part of *The Faerie Queene*. The analysis will then broaden to the wider lexicon of the text and examine the potential use of seemingly irrelevant words that have ties to sixteenth-century Protestant discourse.

In discussing Spenser’s use of words relating to religion and morality, there are three main categories involved: 01.17.04 [Deity], 03.06 [Morality], and 02.03 [Goodness/Badness]. “Deity” encompasses the words relating to overtly religious ideas like God and heaven, whereas “Morality” and “Goodness/Badness” refer to more general concepts often included as a part of, but are not exclusive to, religion such as fairness, faithfulness, goodness, and cruelty (see Table 11). There are other words and categories relating to religion and morality in *The Faerie*

T2/3 Category	Subcategory Heading	Words
Deity	Christian God	God
	Heaven	heaven(s), sky, towers, seat
	Heavenly	heavenly, celestial
Morality	Faithful/ Trustworthy	faithful, true
	Fair/Equitable	fair/fairest
	Morally Foul	foul
Goodness/Badness	Good	good, better, best
	Excellent	lovely, goodly, excellent, admirable, noble, great
	Wretched	wretched, woeful
	Savage, Cruel	cruel

Table 11: Frequencies of words in subcategories of "Deity", "Morality", and "Goodness/Badness"

Queene, but they did not meet the minimum frequency threshold and so will not be included in this discussion.

First, there are some clarifications about the words and categories listed that must be made. An examination of the concordances of "heavenly" demonstrates that the word's most common sense is not "of, in, or belonging to heaven" but rather "divine, sacred, holy" and sometimes "Of more than earthly or human beauty, perfection, or excellence" **(heavenly, adj. and n. 2021)**. While this calls into question the semantic categorisation by the HT Semantic Tagger, the religious connotations of the word in any usage mean that it is still worthy of inclusion. The inclusion of Fair/Equitable is more complicated, however. Many of the uses of "fair" are in phrases like "fair Una" and "where that fair virgin stood; As fair Diana in fresh summers day" (Spenser 1596, ll.5936–7), which suggests they belong in a different category to Fair/Equitable. This categorisation is accurate for some of the senses, but more nuance is needed to separate out the senses that belong elsewhere. Unfortunately, this was unable to be achieved for this project, so the category has been left as is.

First appearances suggest that there are more words relating to aspects of morality than Christianity as there are more subcategories relating to the former than latter. This observation is confirmed when the category frequencies are included: "Deity" has a total of 128 occurrences, "Goodness/Badness" is used 139 times, and "Morality" 154 times (with

"Fair/Equitable" excluded, that figure drops to 44). This indicates that moral virtue is more important in the first book of *The Faerie Queene* than overt references to Christianity. Another way of interpreting this data is comparing the positive and negative words; the categories listed here represent the polar extremes of morality in the same way as the emotion categories. Including "Fair/Equitable", the number of positive categories is 208, or 98 when it is excluded, and negative categories are used 85 times. Given that at least some of the categorisation within "Fair/Equitable" is accurate, positive categories feature more than negative, which is noteworthy as this is the opposite as was found with emotions. A plausible explanation for this disparity can be found when examining the purpose behind suffering for Protestants: when the characters in *The Faerie Queene* experience suffering, they are redeemed and marked as good, which is one of the basic themes of allegory (Tuve 1966, p.49). Therefore, there is not in fact a disparity between the strong presence of suffering and positive moral categories- both ideas work together to reinforce the ideals of Protestantism in the text beyond explicit references.

The influence of religion on texts can go much deeper than overtly religious terms and even references to morality and what constitutes good behaviour. In his examination of the vocabulary used in sixteenth-century religious debate, Smith performed multiple lexical analyses on two corpora – one of evangelical Protestant texts and one of Roman Catholic texts – and established that during the 1560s, Protestant discourse began to break away from other religious discourse and demonstrate its own preference for certain words (2020, p.109). This word list is a useful point of reference for establishing if Spenser's Protestantism permeates the first book of *The Faerie Queene* to such an extent as to still use the hallmarks of Protestant discourse of thirty years prior. Of the twenty-seven words identified by Smith as markers of Protestant discourse, six were not found in *The Faerie Queene* at all. The words that were included are listed in Table 12, along with their frequency in each corpus normalised per 100,000 words.

Word	Faerie Queene Normalised Freq.	Evangelical Normalised Freq.	Roman Catholic Normalised Freq.
blessed	19.7	159.8	51.7
enemy	32.9	34.3	6.2
fire	72.3	24	1.1
friend	24.1	61.6	8.4
glory	26.3	168.5	12.9
grievous	13.1	35.1	3.4
happy	30.7	19.2	3.4
heart	151.1	115	32
heal	8.8	26.4	0.6
hope	33	86.3	9.5
humble	17.5	35.1	3.9
iniquity	2.2	18.4	2.2
joy	61.3	129.4	2.8
light	107.3	65.5	20.8
love	118.3	226.1	23
meek	11	8	0.6
mercy	13	191.7	11.2
neighbour	4.4	91.9	4.5
patience	11	36.7	2.2
righteous	8.8	99.8	1.7
soul	19.7	178.4	18.5

Table 12: Normalised word frequencies from The Faerie Queene and datasets from Smith (2020)

This comparison indicates that the first book of *The Faerie Queene* shares almost half of the evangelical vocabulary to a similar level as the evangelical corpus or compared with the Roman Catholic texts. The words that are overused in the first book of *The Faerie Queene* compared with the evangelical corpus are “fire”, “light”, “heart”, “happy” and “meek”, with “enemy”, “humble”, and “love” being among the words closer to the evangelical corpus in terms of frequency than the Roman Catholic. Given the differences in genre between the reference corpora and *The Faerie Queene* – Smith used sermons and religious treatises as his source material – finding any similarities in vocabulary used is noteworthy. While it could be argued

that some of the words are heavily used due to their necessity for chivalric literature, it is possible that Spenser chose this genre because of its use of such words, making it a useful means for incorporating and disseminating the Protestant style of writing. Whether this is true or not, it is clear that religion is integral to the first book of *The Faerie Queene* at multiple levels of understanding.

6.5 Answers to Research Questions

6.5.1 Which words form the core vocabulary of *The Faerie Queene*?

With regards to the first question posed at the beginning of this research, the twenty most frequent words that make up the core vocabulary of the first book of *The Faerie Queene* are listed in Table 13 alongside the top twenty keywords as identified in WMatrix. Only nouns and adjectives are included in the first list as the most frequently used verbs are mostly auxiliary and modal verbs (see section 5.2.2.3), but the keyword list is not filtered by part of speech. The slight discrepancies in frequencies between the word frequency and keywords lists is due to the more detailed CLAWS tags being present in the word frequency list and not the keyword analysis. In comparing the two, the importance of "knight" to the story is reinforced, since it

Word	Raw Freq.	Word	Raw Freq.	Keyword	LL	Raw Freq.	Keyword	LL	Raw Freq.
knight	169	*sad*	58	*knight*	**388.47**	174	*Duessa*	**127.38**	39
great	125	*blood*	57	*his*	**277.27**	947	*that*	**125.25**	898
fair	108	*lady*	55	*he*	**213.19**	745	*and*	**102.68**	1828
life	72	*death*	55	*her*	**200.65**	639	*fairy*	**97.98**	30
full	70	*hand*	48	*does*	**169.98**	59	*she*	**96.63**	309
heart	69	*eyes*	46	*began*	**169.26**	114	*fair*	**85.87**	130
high	66	*quoth*	46	*him*	**166.56**	390	*wide*	**85.19**	42
man	66	*night*	45	*Una*	**163.31**	50	*eke*	**82.54**	47
day	62	*heaven*	43	*ne*	**135.94**	54	*forth*	**79.87**	92
way	59	*foe*	43	*did*	**132.29**	489	*quoth*	**75.96**	53

Table 13: Core vocabulary ordered by raw frequency and keywords ordered by log-likelihood

tops both the lists. The other words found in both lists are “fair” and “quoth”, with “fairy” and the characters “Una” and “Duessa” appearing among the closed class words in the keyness list. These character names, along with “knight” and “fairy” are examples of text-centred keywords as they are essential for effectively conveying the story. “Fair” could be viewed as a reader-centred keyword as it relates more to the themes within *The Faerie Queene,* but its epithetic usage in phrases like “fair Una” almost make it as text-centric as the character name itself. Nevertheless, the difference in words featured in the two lists indicates that frequency and keyness do not necessarily go hand in hand, and the importance of Una and Duessa is clear despite their lower raw frequency when compared with “knight”.

In terms of raw word frequency, “knight” tops the list as a reference to the many different knights in the text, including the main character the Redcrosse Knight, and a lot of the other words fit in with ideas in chivalric literature: “great”, “fair”, “lady”, “death”, and “foe” in particular. “Hand” and “eyes” are interesting inclusions in this list and refer to the high presence of body part words in the text, and “night” and “day” are used in developing the setting of encounters within the story. What is most interesting about this list, however, is that “heaven” is the only word featured that has overtly religious connotations, which does not represent the importance of religion in the first book of *The Faerie Queene*. Furthermore, as discussed earlier in the chapter, women play an important role in *The Faerie Queene* and the variety of vocabulary used to refer to women obscures this fact in the data by reducing the frequency of each feminine word. Therefore, the fact that “lady” has a lower frequency than both “man” and “knight” does not give the full picture about the position of women in the first book of *The Faerie Queene.*

6.5.2 Does this core vocabulary correspond with the core semantic categories?

Overall, the majority of the semantic categories that the most common words belong to are included in the list of core semantic categories (see table 14). Others are included, such as 02.06.01-03.03.01 [his/her], as they are tagged as open-class categories but are integral to English syntax. There are also some changes in position when compared with the core vocabulary; 01.02.03.23.02 [Heart] and 01.17.04.06 [Heaven] have the same frequency, whereas “heart” occurs twenty-six more times than “heaven”. By referring to semantic categories rather than words, the issue of synonyms is overcome, as demonstrated with words

Category	Freq.	Category	Freq.
02.06.01-03.03.01 [his/her]	1465	*01.16.02.01-10.05.01 [each/every]*	81
01.11.01.07 [Be/remain in specific state/condition]	723	*01.14.05-08.03 [Along which a thing may pass]*	80
01.16.07.02 [All/the whole amount of]	323	*01.12.05.07 [High in position]*	75
01.15.02 [Act/do]	228	*01.02 [Life]*	72
02.06.01-03.03.03 [their]	192	*01.12.04.03.01-01 [full]*	72
03.01.06.01.02.09 [Knight]	174	*01.16.06.03.01-06 [very]*	72
02.02.08-06 [of high/great importance]	149	*01.02.03.23.02 [Heart]*	71
02.01.13.08.09 [Expressing possibility]	148	*01.17.04.06 [Heaven]*	71
01.09.08.03 [See]	126	*02.04.13 [Love]*	70
03.06.03-05 [fair/equitable]	110	*03.10.01.04 [Direct one's course]*	69
02.07.03-18 [begin to speak]	108	*01.02.03.08.04.03.01 [Hand]*	68
01.13.08.03 [Present]	106	*01.15.10.02.01 [Find/discover]*	61
01.04.04.01 [Man]	104	*01.02.04 [Death]*	60
02.06-05 [Have/possess]	97	*01.13.08.04.04 [Old]*	57
02.07.03 [Speak/say/utter]	92	*01.02.03.08.01.04.06 [Eye/eyes]*	56

Table 14: Core semantic categories

relating to women, and the frequencies of semantic categories become more representative of frequency of ideas.

The list of core semantic categories is shaken up, however, when the view is broadened to focus on higher tier categories. It is here that the significance of body part words becomes clear (with 513 occurrences, “The Body” has the third highest tier three category frequency). “Sight/Vision” and “Social Class” also have high frequencies, with 297 and 254 occurrences respectively. Broadening further still to examine the tier two categories is where it becomes clear that emotions play a major role in the first book of *The Faerie Queene*. “Emotion” has a frequency of 473, but there are no individual emotions that have a high enough frequency to be included in either of the core lists with the exception of “sad”. This demonstrates the

benefits of incorporating broader semantic categories into the discussion of the core semantic categories of *The Faerie Queene*.

6.5.3 What can this semantic distribution tell us about Edmund Spenser and his interests?

What is clear from the examination of the core semantic categories is that they feed into the key themes of the first book of *The Faerie Queene* but are not direct representations of them. For example, the categories "Social Class" and "Love" are all important aspects of chivalry, the genre of the text, but the Historical Thesaurus does not have a category or group of categories named "Chivalry"- interpretations must still be made on the part of the researcher as to which categories can be classed as relevant to chivalry. Spenser is clearly interested in the genre of chivalric literature since he wrote *The Faerie Queene* in an updated version of it, and the high frequency of categories like "Social Class", "Death", and "Love" demonstrate that he is remaining faithful to the traditional characteristics of chivalric literature while altering the genre enough to fit the allegorical messages included within.

However, other more concrete conclusions about Spenser's interests can still be made. "Deity", "Morality" and the suffering aspect of "Emotion" contribute to Spenser's exploration of Protestantism in the text, which must reflect his own views on what constitutes a good Protestant. Furthermore, gender roles and female agency are treated differently in the first book of *The Faerie Queene* when compared to other texts of the period. Spenser also uses body part terms as a way of grounding abstract concepts like love and ancestry, in addition to furthering character description. These notions have all been explored through literary analysis, but the application of corpus analysis techniques and the use of the Historical Thesaurus means that these observations can now be corroborated with data in a new and exciting way.

7 CONCLUSION

This research aimed to use lexical frequency and semantic category analyses to uncover the interests of Edmund Spenser hidden in the first book of *The Faerie Queene*. The core vocabulary was identified by using WMatrix and examined in detail using AntConc, with a reference corpus of thematically similar period texts providing a point of comparison for establishing keyness and frequency significance. For identifying the core semantic categories, the text was tagged according to the hierarchy of the Historical Thesaurus of English and the resulting data was compared with a representation of the HT data itself.

One of the main achievements of this research was bringing together multiple domains, including corpus stylistics, to create a well-rounded viewpoint of Spenser and his outlooks as indicated by the first book of *The Faerie Queene*. The body of work on the text is extensive, but this research explored the text in a new way; the chosen methodology was successful in identifying the text's core vocabulary and semantic categories, and in-depth examination of the resultant data revealed Spenser's interests in religion and chivalry hidden among a dearth of varied literary techniques. Incorporating ideas about these key themes in the text explored in previous literary analyses acted as a guide for focusing this research, and the results corroborate the findings of scholarly work on *The Faerie Queene*.

In addition to providing an alternative viewpoint on the first book of *The Faerie Queene* through quantitative analysis, this research functioned as a pilot study for using the HT in this type of investigation with the aim of establishing its potential for being used in future research. Given the volume and quality of data provided by the lexical and semantic category analyses, the methodology has proven to be effective in providing the basis for more data-driven examinations of literary works.

There are a number of ways in which the research undertaken here could be expanded and built upon. The analysis of the first book of *The Faerie Queene* could be expanded to examine text- and reader-centred keywords in greater detail, which ties into Spenser's use of abstract and tangible sematic categories. There are also multiple areas touched on in this thesis that are worthy of greatly expanded research: the comparative treatment of male and female characters; Spenser's treatment of emotions, especially negative ones; and how his use of

language aligns him with other Protestant writers, to name but a few. Furthermore, instead of focusing on the first book of *The Faerie Queene*, all six books could be included in the analysis. The core vocabulary and themes of the entire text and each individual book could be identified and contrasted to see if Spenser focuses on the same topics throughout the work or not, or whether he treats the same issues in different ways in different stories. Alternatively, *The Faerie Queene* could be contrasted with other Spenserian works, such as the collection of poems *Complaints* or *The Shepheardes Calender*. The benefits of comparing *The Faerie Queene* with the latter are especially strong, since *The Shepheardes Calender* is Spenser's first published work and *The Faerie Queene* his last, so a detailed analysis of the evolution of his style and interests over the course of his life could be performed.

Moving away from Spenser, the methodology could be adapted to place more of a focus on the reference corpus of period texts. If it was adapted to feature more works of literature, then it would be a more solid basis for comparison for *The Faerie Queene*. The reference corpus could also be semantically tagged to provide a more appropriate reference point for relevant semantic categories than the entire Historical Thesaurus as was used in this project, which would allow for a more direct comparison between Spenser's language and sixteenth-century English more broadly. The methodology employed in this research is highly adaptable and its successful use has demonstrated its potential to be used in a variety of contexts in historical corpus linguistics.

8 BIBLIOGRAPHY

8.1 PRIMARY SOURCE

Spenser, E. (1596) The First Book of the Faerie Qveene. Contayning The Legende of the Knight of the Red Crosse, or of Holinesse. In: *The Faerie Queene Disposed into twelue bookes, fashioning XII. morall vertues.* London, William Ponsonbie; Early English Books Online Text Creation Partnership.

8.2 SOFTWARE

Anthony, L. (2019) *AntConc*. Tokyo, Japan, Waseda University.

Baron, A. and Rayson, P. (2008) VARD 2: A tool for dealing with spelling variation in historical corpora. *Postgraduate conference in corpus linguistics*. Aston University, Birmingham.

Piao, S., Dallachy, F., Baron, A., Demmen, J., Wattam, S., Durkin, P., McCracken, J., Rayson, P. and Alexander, M. (2017) A time-sensitive historical thesaurus-based semantic tagger for deep semantic annotation. *Computer Speech and Language*, 46, pp. 113–135.

Plaisant, C. (2003) *Treemap* [Internet]. Available from https://www.cs.umd.edu/hcil/treemap/. [Accessed 7th July 2021].

Rayson, P. (2009) *Wmatrix: a web-based corpus processing environment*. Computing Department, Lancaster University.

8.3 SECONDARY SOURCES

aread | arede | areed, v. (2020) In: *OED Online*. Oxford University Press. Available from www.oed.com/view/Entry/10508.

arm, n.1. (2021) In: *OED Online*. Oxford University Press. Available from www.oed.com/view/Entry/10808. [Accessed 6th July 2021].

arms, n. (2021) In: *OED Online*. Oxford University Press. Available from www.oed.com/view/Entry/10809. [Accessed 6th July 2021].

Alexander, M. (2011) The various forms of civilization arranged in chronological strata: manipulating the HTOED. In: Adams, M. and Imartino, G. eds. *Cunning Passages, Contrived Corridors: Unexpected Essays in the History of Lexicography*, Lexicography Worldwide. Monza, Italy, Polimetrica Press.

Algee-Hewitt, M. (2021) *About* [Internet]. Available from https://litlab.stanford.edu/. [Accessed 9th August 2021].

Allison, S., Gemma, M., Heuser, R., Moretti, F., Tevel, A. and Yamboliev, I. (2013) Style at the Scale of the Sentence. *Pamphlets of the Stanford Literary Lab*, (5).

Altermatt, T.W. (2001) *Chivalry: The relation between a cultural script and stereotypes about women*. Thesis, University of Illinois at Urbana-Champaign.

Anderson, J.H. (2018) Spenser's Narrative Figuration of Women in The Faerie Queene. *Research in Medieval and Early Modern Culture*, 1.

Anon (2020) aread | arede | areed, v. In: *OED Online*. Oxford University Press. Available from www.oed.com/view/Entry/10508.

Anon (2020) dight, v. In: *OED Online*. Oxford University Press. Available from www.oed.com/view/Entry/52604. [Accessed 1st February 2021].

Anon (2021) arm, n.1. In: *OED Online*. Oxford University Press. Available from www.oed.com/view/Entry/10808. [Accessed 6th July 2021].

Anon (2021) arms, n. In: *OED Online*. Oxford University Press. Available from www.oed.com/view/Entry/10809. [Accessed 6th July 2021].

Anon (2021) blood, n. (and int.). In: *OED Online*. Oxford University Press. Available from www.oed.com/view/Entry/20391.

Anon (2021) heavenly, adj. and n. In: *OED Online*. Oxford University Press. Available from www.oed.com/view/Entry/85220. [Accessed 27th July 2021].

Anthony, L. (2019) *AntConc*. Tokyo, Japan, Waseda University.

Archer, D., Kytö, M., Baron, A. and Rayson, P. (2015) Guidelines for normalising Early Modern English corpora: Decisions and justifications. *ICAME Journal*, 39, pp. 5–25.

Baron, A. (2013) *User Guide* [Internet]. Available from http://ucrel.lancs.ac.uk/vard/userguide/. [Accessed 27th January 2021].

Baron, A. and Rayson, P. (2008) VARD 2: A tool for dealing with spelling variation in historical corpora. Postgraduate conference in corpus linguistics. Aston University, Birmingham.

Beal, P. (2013) *Edmund Spenser* [Internet]. Available from https://celm-ms.org.uk/introductions/SpenserEdmund.html. [Accessed 8th February 2021].

Berber-Sardinha, T. (2000) Comparing corpora with WordSmith Tools: How large must the reference corpus be? *The Workshop on Comparing Corpora*, pp. 7–13.

Borris, K. (2000) Flesh, Spirit, and the Glorified Body: Spenser's Anthropomorphic Houses of Pride, Holiness, and Temperance. *Spenser Studies*, 15 (1), pp. 17–52.

Brown, R.D. (2019) Wise wights in privy places: Rhyme and stanza form in Spenser and Chaucer. In: *Rereading Chaucer and Spenser*. Manchester, Manchester University Press, pp. 113–136.

Brown, R.D. and Lethbridge, J.B. (2021) *A Concordance to the Rhymes of The Faerie Queene*. Manchester, Manchester University Press.

Canny, N. (1983) Edmund Spenser and the Development of an Anglo-Irish Identity. *The Yearbook of English Studies*, 13, pp. 1–19.

Carroll, S. (2018) *Medievalism in A Song of Ice and Fire and Game of Thrones*. Cambridge, D. S. Brewer.

Craig, J. (1972) The Image of Mortality: Myth and History in the Faerie Queene. *ELH*, 39 (4), pp. 520–544.

Culpeper, J. (2009) Keyness: Words, Parts-of-Speech and Semantic Categories in the Character-Talk of Shakespeare's Romeo and Juliet. *International Journal of Corpus Linguistics*, 14 (1), pp. 29–59.

Davis, B.E.C. (1933) *Edmund Spenser: A Critical Study*. Cambridge, Cambridge University Press.

Davis, W. (2002) Spenser and the History of Allegory. *English Literary Renaissance*, 32 (1), pp. 152–167.

Friedman, Y. (2019) Masculine attributes of the other: The shared knightly model. In: *Crusading and Masculinities*. Routledge, pp. 89–99.

Gans, N.A. (1979) Archaism and Neologism in Spenser's Diction. *Modern Philology*, 76 (4), pp. 377–9.

Green, P.D. (1974) Spenser and the Masses: Social commentry in the Faerie Queene. *Journal of the History of Ideas*, 35 (3), pp. 389–406.

Greenblatt, S. ed. (2018) *The Norton Anthology of English Literature: The Sixteenth Century and The Early Seventeenth Century*. New York; London, W. W. Norton & Company.

Hadfield, A. (2011) Spenser And Religion—Yet Again. *Studies in English Literature, 1500-1900*, 51 (1), pp. 21–46.

Hadfield, A. (2014) *Edmund Spenser: A Life*. New York, Oxford University Press.

Hadfield, A. (2021) *Spenser, Edmund* [Internet]. Available from https://doi.org/10.1093/ref:odnb/26145. [Accessed 13th July 2021].

Hamilton, A.C. (2019) Our new poet: Spenser,'well of English undefyld'. In: *A Theatre for Spenserians*. Toronto, University of Toronto Press, pp. 101–123.

Hardie, A. (2014). Log ratio: An informal introduction. *ESRC Centre for Corpus Approaches to Social Science (CASS)*.

Heuser, R. and Le-Khac, L. (2012) A Quantitative Literary History of 2,958 Nineteenth-Century British Novels: The Semantic Cohort Method. *Pamphlets of the Stanford Literary Lab*, (4).

Heuser, R., Moretti, F. and Steiner, E. (2016) The Emotions of London. *Pamphlets of the Stanford Literary Lab*, (13).

Heuser, R. and Pestre, D. (2015) Bankspeak. The language of World Bank Reports, 1946-2012. *Pamphlets of the Stanford Literary Lab*, (9).

Higgins, A. (1990) Spenser Reading Chaucer: Another Look at the 'Faerie Queene' Allusions. *The Journal of English and Germanic Philology*, 89 (1), pp. 17–36.

Hume, A. (1984) *Edmund Spenser: Protestant Poet*. Cambridge, Cambridge University Press.

Keen, M.H. (2005) *Chivalry*. New Haven and London, Yale University Press.

King, A. (2010) Spenser, Chaucer, and Medieval Romance. In: *The Oxford Handbook of Edmund Spenser*. Oxford, Oxford University Press, pp. 554–572.

King, J.N. (1990) *Spenser's Poetry and the Reformation Tradition*. Princeton, Princeton University Press.

Knuuttila, S. (2004) *Emotions in Ancient and Medieval Philosophy*. Oxford, Clarendon Press.

Knuuttila, S. (2012) Sixteenth-Century Discussions of the Passions of the Will. In: Pickavé, M. and Shapiro, L. eds. *Emotion and Cognitive Life in Medieval and Early Modern Philosophy*. Oxford, Clarendon Press, pp. 116–132.

Leech, G.N., Garside, R. and Bryant, M. (1994) CLAWS4: The tagging of the British National Corpus. In: *Proceedings of the 15th International Conference on Computational Linguistics (COLING 94)*. Kyoto, Japan, pp. 622–628.

Mahlberg, M. (2007a) Clusters, key clusters and local textual functions in Dickens. *Corpora*, 2 (1), pp. 1–31.

Mahlberg, M. (2007b) Corpus stylistics: bridging the gap between linguistic and literary studies. In: *Text, Discource and Corpora: Theory and Analysis*. London and New York, Continuum, pp. 219–246.

Mahlberg, M. (2010) A corpus stylistic perspective on Dickens' Great Expectations. In: Lambrou, M. and Stockwell, P. eds. *Contemporary Stylistics*. London, Continuum, pp. 19–31.

Mahlberg, M. and McIntyre, D. (2011) A case for corpus stylistics: Ian Fleming's Casino Royale. *English Text Construction*, 4 (2), pp. 204–227.

Maley, W. (1994a) *A Spenser Chronology*. London, Palgrave Macmillan.

Maley, W. (1994b) Spenser's Irish English: Language and Identity in Early Modern Ireland. *Studies in Philology*, 91 (4), pp. 417–431.

Mallette, R. (1997) *Spenser and the Discourses of Reformation England*. Lincoln, University of Nebraska Press.

McElderry, B.Ro. (1932) Archaism and innovation in Spenser's poetic diction. *PMLA*, 47 (1), pp. 144–170.

Moulton, I. (2014) *Love in Print in the Sixteenth Century: The Popularization of Romance*. Basingstoke, UK, Palgrave Macmillan.

Narveson, K. (1999) Flesh, Excrement, Humors, Nothing: The Body in Early Stuart Devotional Discourse. *Studies in Philology*, 96 (3), pp. 313–333.

Osgood, C.G. (1915) *A concordance to the poems of Edmund Spenser*. Washington, D.C., Carnegie Institute of Washington.

Padelford, F.M. (1941) Aspects of Spenser's Vocabulary. In: Maxwell, B., Briggs, W.D., Johnson, F.R. and Thompson, E.N.S. eds. *Renaissance Studies in Honor of Hardin Craig*, special issue 20. Stanford, Stanford University Press, pp. 279–83.

Parker, R.E. (1925) Spenser's Language and the Pastoral Tradition. *Language*, 1 (3), pp. 80–87.

Piao, S., Dallachy, F., Baron, A., Demmen, J., Wattam, S., Durkin, P., McCracken, J., Rayson, P. and Alexander, M. (2017) A time-sensitive historical thesaurus-based semantic tagger for deep semantic annotation. *Computer Speech and Language*, 46, pp. 113–135.

Plaisant, C. (2003) *Treemap* [Internet]. Available from https://www.cs.umd.edu/hcil/treemap/. [Accessed 7th July 2021].

Quilligan, M. (2018) *The Language of Allegory*. Ithaca, NY, Cornell University Press.

Rayson, P., Archer, D., Piao, S. and McEnery, T. (2004) The UCREL Semantic Analysis System.

Rayson, P., Archer, D. and Smith, N. (2005) VARD versus Word: A comparison of the UCREL variant detector and modern spell checkers on English historical corpora. In: *Proceedings of the Corpus Linguistics 2005 conference*. Birmingham, UK.

Rayson, P., Emmet, L., Garside, R. and Sawyer, P. (2000) The REVERE Project: Experiments with the Application of Probabilistic NLP to Systems Engineering. In: *International Conference on Application of Natural Language to Information Systems*. Berlin, Heidelberg, pp. 288–300.

Ryrie, A. (2013) *Being Protestant in Reformation Britain*. Oxford, Oxford University Press.

Sandberg, J. (2016) *Flesh becoming word: Eucharist and allegory in early modern England*. PhD Thesis, Southern Methodist University.

Smith, J.J. (2020) Godly vocabulary in Early Modern English religious debate. In: Jonsson, E. and Larsson, T. eds. *Voices past and present - Studies of involved, speech-related and spoken texts: In honor of Merja Kytö*, Studies in Corpus Linguistics. Amsterdam, John Benjamins Publishing, pp. 95–112.

Spenser, E. (1932) *THE FAERIE QVEENE. Disposed into Twelue Bookes, Fashioning XII. Morall Vertues*. Baltimore, John Hopkins Press.

Stephens, D. (2010) Spenser's Language(s): Linguistic Theory and Poetic Diction. In: McCabe, R.A. ed. *The Oxford Handbook of Edmund Spenser*. Oxford, Oxford University Press, pp. 367–384.

Stubbs, M. (1995) Collocations and semantic profiles: On the cause of the trouble with quantitative studies. *Functions of Language*, 2 (1), pp. 23–55.

Turnage, M. and Spenser, E. (1970) Samuel Johnson's Criticism of the Works of Edmund Spenser. *Studies in English Literature, 1500-1900*, 10 (3), pp. 557–567.

Tuve, R. (1966) *Allegorical Imagery: Some Medieval Books and Their Posterity*. Princeton, Princeton UP.

Walls, K. (2020) The Christian Lark: Spenser's Faerie Queene I. x. 51 and Shakespeare's Sonnet 29. *Explorations in Renaissance Culture*, 46 (2), pp. 200–218.

Wauchope, G.A. ed. (1903) *Spenser's The Faerie Queene, Book I*. London, MacMillan & Co.

Webster, J. (1976) Oral Form and Written Craft in Spenser's Faerie Queene. *Studies in English Literature, 1500-1900*, 16 (1), pp. 75–93.

Yamashita, H. ed. (1993) *A Textual Companion to The Faerie Queene, 1590*. Tokyo, Kenyusha.

Yamashita, H., Matsuo, M., Suzuki, T. and Sato, H. eds. (1990) *A Comprehensive Concordance to The Faerie Qveene 1590*. Tokyo, Kenyusha.

9 APPENDIX A: WMATRIX REFERENCE CORPUS

This reference corpus was created to provide a point of comparison for establishing lexical keyness in the first book of *The Faerie Queene*. WMatrix automatically includes modern corpora like those derived from the British National Corpus (BNC), but given that *The Faerie Queene* is written in Early Modern English, a more suitable comparison corpus needed to be created. The details of the EEBO texts included in this corpus are listed below.

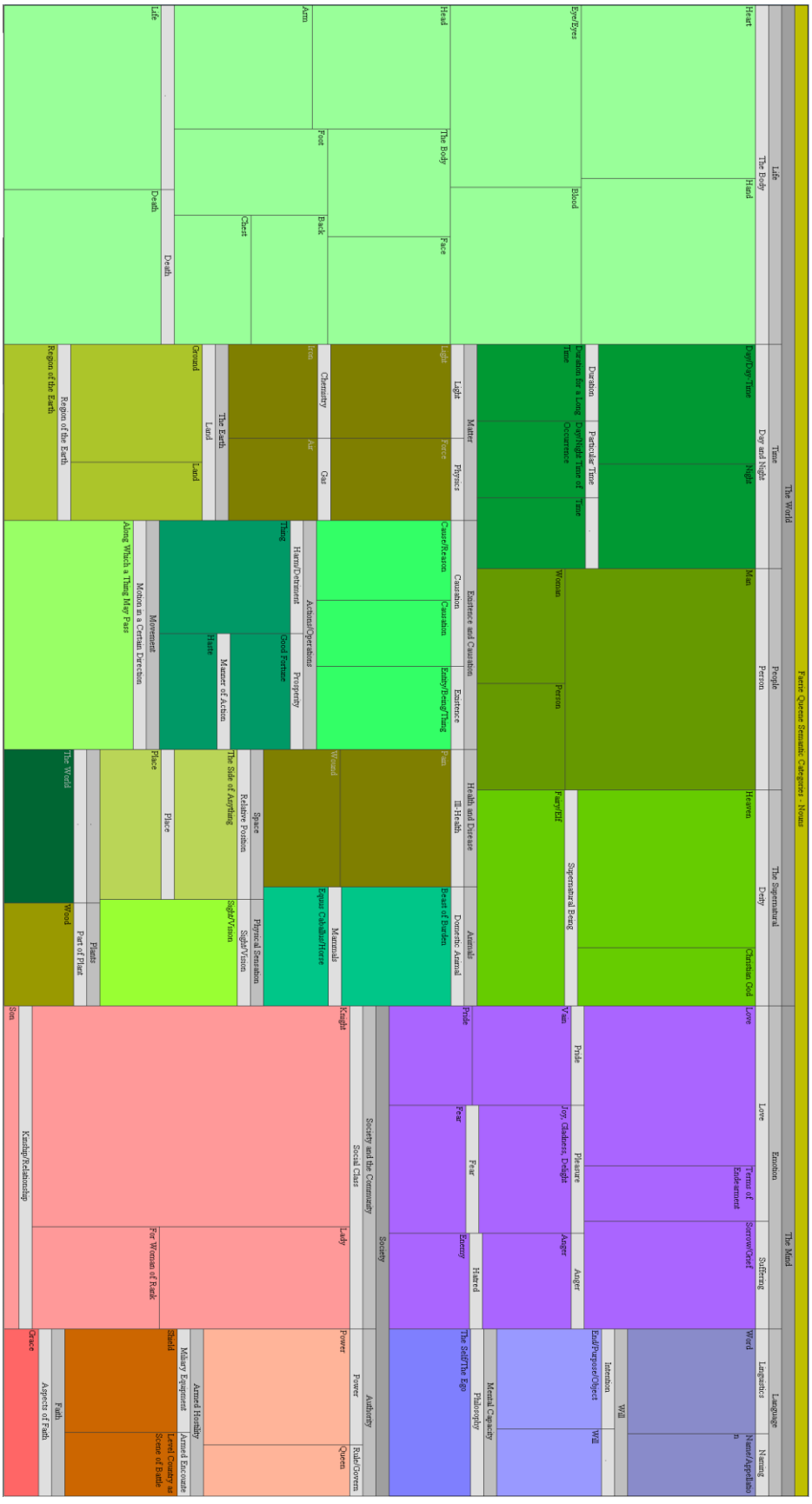
Author	Title	Date	Pages
Huggarde, Miles.	A mirrour of loue, which such light doth giue, that all men may learne, how to loue and liue. Compiled and set furth by Myles Hogarde seruaunt to the quenes highnesse	1555	60
Fisher, John	[3 dialogues in verse, between Gelasimus and Spudaeus, Eda and Agna, and Wisdom and Wyll]	1558	54
Holme, Wilfrid.	The fall and euill successe of rebellion from time to time wherein is contained matter, moste meete for all estates to vewe. Written in old Englishe verse, by VVilfride Holme.	1572	70
Hake, Edward	A commemoration of the most prosperous and peaceable raigne of our gratious and deere soueraigne lady Elizabeth by the grace of God of England, Fraunce and Irelande, Queene &c. Now newly set foorth this. xvii. day of Nouember, beyng the first day of the. xviii. yeere of her Maiesties sayd raigne. By Edw. Hake. Gent.	1575	40
Boorde, Andrew; More, Thomas, Sir	A ryght pleasaunt and merye historie, of the mylner of Abyngton, with his wife, and his fayre daughter: and of two poore scholers of Cambridge Wherevnto is adioyned another merye Iest, of a sargeaunt that woulde haue learned to be a fryar.	1576	30
Dallington, Robert	A booke of epitaphes made vpon the death of the right worshipfull Sir VVilliam Buttes knight vvho deceased the third day of September, anno 1583.	1583	57
Aske, James	Elizabetha triumphans Conteyning the dammed practizes, that the diuelish popes of Rome haue vsed euer sithence her Highnesse first comming to the Crowne, by mouing her wicked and traiterous subjects to rebellion and conspiracies, thereby to bereaue her Maiestie both of her lawfull seate, and happy life. VVith a declaration of the manner how her excellency was entertained by her souldyers into her campe royall at Tilbery in Essex: and of the ouerthrow had against the Spanish fleete: briefly, truly, and effectually set foorth. Declared, and handled by I.A.	1588	44

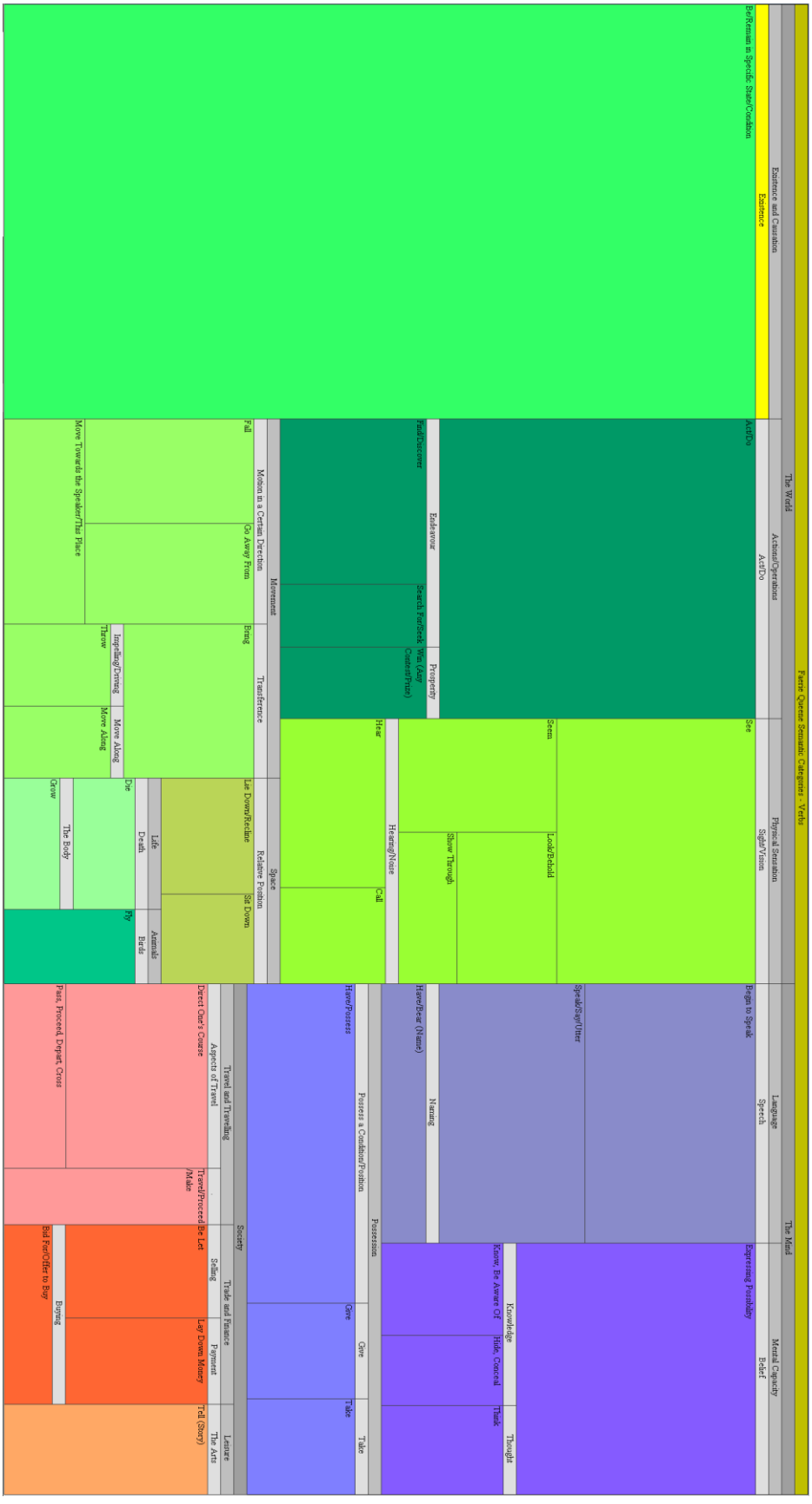
Churchyard, Thomas	A handeful of gladsome verses, giuen to the Queenes Maiesty at Woodstocke this prograce. 1592. By Thomas Churchyarde	1592	22
Drayton, Michael	Matilda The faire and chaste daughter of the Lord Robert Fitzwater. The true glorie of the noble house of Sussex.	1594	64
Jones, Richard; Breton, Nicholas	The arbor of amorous deuises VVherin, young gentlemen may reade many plesant fancies, and fine deuises: and thereon, meditate diuers sweete conceites, to court the loue of faire ladies and gentlewomen by N.B. Gent.	1597	52
Rogers, Thomas	Celestiall elegies of the goddesses and the Muses de-deploring [sic] the death of the right honourable and vertuous ladie the Ladie Fraunces Countesse of Hertford, late wife vnto the right honorable Edvvard Seymor Vicount Beauchamp and Earle of Hertford. Whereunto are annexed some funerall verses touching the death of Mathevv Evvens Esquire, late one of the barons of her Maiesties Court of Eschequer, vnto whome the author hereof was allyed. By Thomas Rogers Esqui[r]e.	1598	64
Davies, John, Sir	Hymnes of Astraea in acrosticke verse.	1599	30
Norden, John	A pensiue soules delight The contents whereof, is shewen in these verses following. I. The pensiue soule recounteth in this place, Elizaes troubles, and Elizaes grace. 2. Here are expressed the stratagems of foes, Elizaes conquests, and their falls that rose. 3. Here is set forth Elizaes lenitie, and locust-Catholickes superbitie. By Iohn Norden.	1603	48
Scoloker, Anthony; Raleigh, Walter, Sir	Daiphantus, or the passions of loue Comicall to reade, but tragicall to act: as full of wit, as experience. By An. Sc. gentleman. Wherevnto is added, The passionate mans pilgrimage.	1604	52
Niccols, Richard	The cuckovv. Richardus Niccols, in Artibus Bac. Oxon	1607	56
Scott, Thomas	The second part of Philomythie, or Philomythologie Containing certaine tales of true libertie. False friendship. Power vnited. Faction and ambition. / By Thomas Scot Gent..; Philomythie. Part 2.	1616	40
Colville, Elizabeth Melvill, Lady Colville of Culros	A godlie dreame compiled by Elizabeth Melvill, Ladie Culros younger, at the request of a friend.	1620	23
Holland, Hugh	A cypres garland For the sacred forehead of our late soueraigne King Iames. / By Hugh Holland.	1625	25
Hodson, William	The plurisie of sorrovv, let blood in the eye-veine: or, The Muses teares for the death of our late soueraigne,	1625	31

	Iames King of England, &c. By Will. Hogson Mag: in Art: Cantab		
Hamilton, Francis, of Silvertown-hill	King Iames his encomiumae· Or A poeme, in memorie and commendation of the high and mightie monarch Iames; King of great Britaine. France, and Ireland &c. our late soveraigne, who deceased at Theobalds. vpon Sunday the 27. of March. 1625. By Francis Hamiltoun, of Silvertown-hill	1626	34
Taylor, John	A dog of vvar, or, The trauels of Drunkard, the famous curre of the Round-Woolstaple in Westminster His seruices in the Netherlands, and lately in France, with his home returne. By Iohn Taylor. The argument and contents of this discourse is in the next page or leafe.	1628	47
Drummond, William; Forbes, Walter,	The entertainment of the high and mighty monarch Charles King of Great Britaine, France, and Ireland, into his auncient and royall city of Edinburgh, the fifteenth of Iune, 1633	1633	40
Anon	The pleasant history of Cawwood the rooke. Or, The assembly of birds with the severall speeches which the birds made to the eagle, in hope to have the government in his absence: and lastly, how the rooke was banished; with the reason why crafty fellowes are called rookes. As also fit morralls and expositions added to every chapter.	1640	26
Howell, James	Parables reflecting upon the times	1643	18
Quarles, Francis	Hosanna, or, Divine poems on the passion of Christ by Francis Quareles.	1647	40

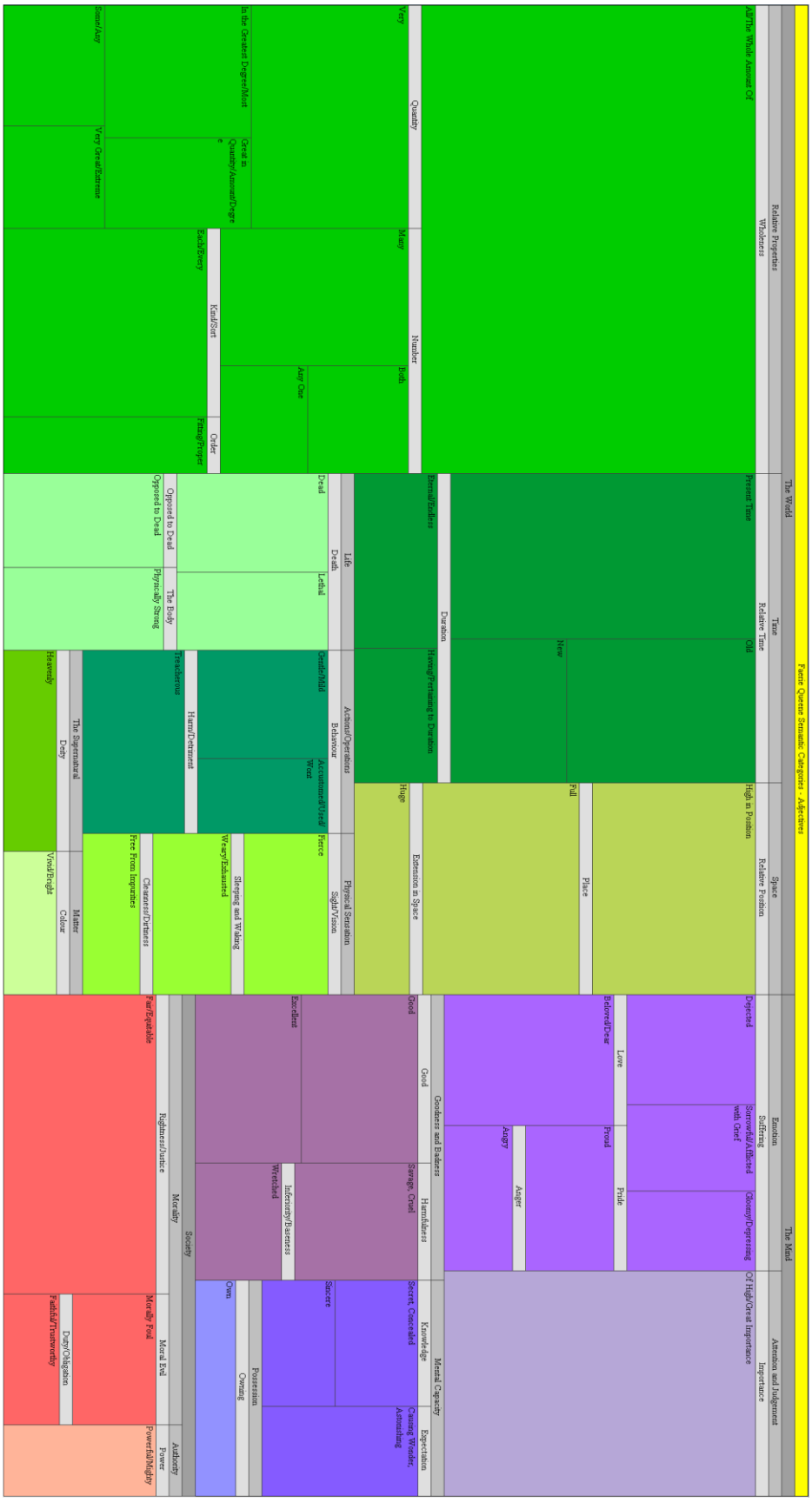
10 Appendix B: Specific Part-of-Speech Treemaps

Treemap of Noun Categories





Treemap of Verb Categories



Treemap of Adjective Categories