Historicising corpus stylistics:
keywords and collocation in William Langland's *Piers Plowman*

Jacqueline Cordell

This thesis is submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy in Applied Linguistics

Department of English Language and Applied Linguistics
University of Nottingham
September 2018

THESIS CONTAINS CD

Abstract

This study presents a uniquely interdisciplinary survey of the language in *Piers Plowman*. Drawing from digital methods and theories developed in corpus linguistics, this investigation examines the relationship between linguistic form and literary function in the poem's lexis by comparing the language contained in the text to a reference corpus of contemporary late medieval narrative poetry. This comparison reveals quantitatively-established text features and patterns that, when placed within theoretical frameworks in literary linguistics, offer a point of departure for analysing aspects of the poem's literary style with a pivotal focus on lexis addressing themes of consumption and the law. This approach does not provide a comprehensive analysis of the linguistic features contributing to the style characterising *Piers Plowman,* but rather operates as a case study illustrating the analytical benefits of taking a corpus approach to the poem.

This research follows an existing tradition of corpus linguistic investigation of literary style but is groundbreaking in being the first to examine pre-Shakespearian texts written in Middle English. It builds on existing corpus linguistic theory and methods, testing their applicability to (and durability in) much more distant historical fields than generally explored in corpus stylistic investigations. Extending corpus analysis of style from early modern to late medieval English literature answers calls for further examination of style in historical texts (Short 2008: 3; Busse 2010b: 33, Bray 2014: 485), and contributes to existing frameworks and methods aimed at studying large amounts of literary discourse embedded within its appropriate historical context. In doing so, my investigation also offers an innovative exploratory approach to the study of poetry within a historical dimension.

Acknowledgements

Many friends and colleagues have assisted me in the completion of this thesis. I would firstly like to thank the University of Nottingham for the funding that enabled the further development of the ideas conceived in my Masters study into my current doctoral research. I am especially indebited to my supervisors Mike Jones, Kevin Harvey, and Michaela Mahlberg without whose collective guidance, encouragement, insight, patience, and support this research would not have been possible. I would like to thank Peter Stockwell for his invaluable feedback at various stages in the research process, and also Monica Chiu and Cord Whitaker for believing in me all those years ago. Finally, my deepest gratitude goes to my family and friends for their constant and unconditional support, and to Greg Newstead for all the cat videos.

Table of contents

Abstract …… 3
Acknowledgements …… 4
List of figures …… 8
List of tables …… 10
List of appendices …… 11

1. Introduction …… 12
    1.1 Why study Piers Plowman? …… 12
    1.2 Research questions …… 15
    1.2 Thesis structure …… 17

2. Linguistic approaches to historical style: examining Piers Plowman through corpus analysis …… 21
    2.1 Piers Plowman …… 21
        2.1.1 The text …… 23
        2.1.2 Features of the text …… 26
        2.1.3 Versions of the text …… 29
        2.1.4 Modern reception …… 31
        2.1.5 Critical literary scholarship …… 33
        2.1.6 Themes in Piers Plowman: consumption and the law 38
            2.1.6.1 Consumption …… 38
            2.1.6.2 Legality and the law …… 42
    2.2 Literary linguistics …… 49
        2.2.1 What is literary style? …… 51
        2.2.2 (New) historical stylistics …… 59
        2.2.3 Style in Middle English poetry …… 63
    2.3 A corpus approach to style …… 67
        2.3.1 Corpus stylistics …… 70
        2.3.2 Taking a CS approach to Piers Plowman …… 77
        2.3.3 Keywords …… 80
        2.3.4 Collocations …… 88
    2.4 Studying historical language using corpora: some considerations 92
        2.4.1 Spelling variation …… 94
        2.4.2 Orthographic normalisation …… 98

3. Methodology …… 106
    3.1 Building the corpora …… 108

3.1.1 The corpora ……. 108
3.1.2 Corpus compilation …… 110
3.1.2.1 PPCorp …….. 110
3.1.2.2 MEPoetLitCorp …… 113
3.2 Annotating the corpora …….. 120
3.2.1 Manual annotation …… 121
3.2.2 Automated annotation …… 126
3.3 Impact of regularisation ……. 128
3.3.1 Benchmarks of effectiveness …… 130
3.3.1.1 Type reduction …… 131
3.3.1.2 Variant replacement …….. 133
3.3.1.3 Amounts of modernisation …….. 135
3.3.2 Conclusions ……. 143
3.4 Data analysis ……. 144
3.4.1 Word frequencies and KWs …….. 145
3.4.2 KW selection ……. 152
3.4.3 Collocational analysis …….. 153
3.4.4 Contributions to literary scholarship ……. 156

4. Surveying the corpus: frequencies and keywords in *PPCorp* ……. 158
4.1 Word frequencies ……. 158
4.2 Keywords and themes ……. 168
4.3 From corpus survey to textual analysis …… 175

5. Lexical representation of consumption ……. 177
5.1 Keywords ……. 178
5.2 KW classification by text function …… 181
5.3 Dual functionality via property allegory: *hunger* and *glutton* 185
5.3.1 Collocations …….. 187
5.3.2 Lexical priming …… 194
5.3.3 Degrees of lexical priming …….. 197
5.3.3.1 Animacy …… 197
5.3.3.2 Foregrounding and commonality …… 202
5.3.4 Hunger and glutton in the literary criticism 204
5.3.5 PPCorp vs. MEPoetLitCorp …….. 207
5.4 Dual functionality in fictional world development: *drink* 210
5.4.1 Collocations …….. 211
5.4.2 Evaluative language and consumption ……. 218
5.4.3 Types of appraisal …….. 227

5.4.3.1 Affect 231
5.4.3.2 Appreciation 232
5.4.3.3 Judgement 233
5.4.4 Judging drink(ing) through propriety 236
5.4.4.1 Social praise 239
5.4.4.2 Social condemnation 240
5.4.5 PPCorp vs. MEPoetLitCorp 242
5.4.5.1 MVCorp 243
5.4.5.2 ARCorp 248
5.5 Conclusion 252

6. Lexical representation of the law 255
6.1 Keywords 256
6.2 KW classification by text function 258
6.3 Dual functionality via class allegory: *doctor* and *justice* 266
6.3.1 Collocations 267
6.3.2 Lexical priming 272
6.3.2.1 Definiteness 274
6.3.3 Allegory in the literary criticism 279
6.3.4 PPCorp vs. MEPoetLitCorp 283
6.4 Multi-functionality in world-building and thematic signalling: *reason* 286
6.4.1 Collocations 287
6.4.2 Technicality and legal language 294
6.4.2.1 Technical KWs 297
6.4.2.2 General KWs with legal applications 302
6.4.2.3 Sub technical KWs 305
6.4.3 PPCorp vs. MEPoetLitCorp 312
6.4 Conclusion 323

7. Conclusion 326
6.1 Main findings 326
6.2 Limitations 331
6.3 Future directions 334

References 337
Appendix A 364

List of figures

Figure 2.1: The ABC cycle of the *PP* manuscripts …… 30
Figure 2.2: Corpus stylistic circle …… 54
Figure 2.3: Approaches to literary style …… 58
Figure 2.4: Concordances in a medieval critical context …… 76
Figure 2.6: Corpus-generated keyword concordances …… 86
Figure 2.7: Corpus view of collocational patterns …… 90
Figure 2.9: Problems with spelling variation in word lists …… 97
Figure 2.10: Variation in ME versus EModE …… 102
Figure 3.6: Manually annotating *PPCorp* using *VARD* …… 124
Figure 3.7: The annotation process …… 127
Figure 3.8: VARDed versus non-VARDed ME text …… 129
Figure 3.10: Pre- vs. post-annotation type amounts …… 133
Figure 3.13: Batch distribution in first manual training stage …… 139
Figure 3.14: Texts directly trained on VARD …… 140
Figure 5.2: Functional classification of the consumption KWs …… 182
Figure 5.5: Concordances lines for *hunger* in *PPCorp* …… 192
Figure 5.6: Concordance lines for *glutton* in *PPCorp* …… 193
Figure 5.7: Creation of text function in property allegory …… 196
Figure 5.8: Hierarchical scale of animacy …… 199
Figure 5.10: Approaches to text function in the literary criticism 205
Figure 5.12: *Food and drink* in *PPCorp* …… 213
Figure 5.13: Concordance lines for *drink* in *PPCorp* …… 214
Figure 5.14: Martin and White's (2005) framework for linguistic appraisal 223
Figure 5.16: Examples of evaluative polarity …… 226
Figure 5.20: Concordance lines for *drink* in *ARCorp* …… 249
Figure 6.2: Functional classification of the legal KWs …… 260
Figure 6.3: Concordance lines for *pena* in *PPCorp* …… 263
Figure 6.6: Concordance lines for *justice* in *PPCorp* …… 271
Figure 6.7: Concordance lines for *doctor* in *PPCorp* …… 271
Figure 6.8: Creation of text function in class allegory …… 276
Figure 6.9: Updated animacy scale, including definiteness …… 280
Figure 6.11: Concordance lines for *doctor* in *MEPoetLitCorp* …… 284
Figure 6.13: Concordance lines for *reason* in *PPCorp* 290
Figure 6.15: Concordance lines for *perpetual* in *PPCorp* 303
Figure 6.17: Lexical consellation of *reason* in *MEPoetLitCorp* 314
Figure 6.18: Lexical constellation of *reason* in *PPCorp* 317
Figure 6.19: *Reason* binomials in *MEPoetLitCorp* 322

Figure 7.1: Functional classification of property versus class allegory in PP using Mahlberg & McIntyre's (2011 framework) ……. 330

List of tables

Table 2.5: Levels of stylistic deviation …… 78
Table 2.8: Levels of extended lexical meaning …… 91
Table 3.1: Study's target and reference corpora …… 110
Table 3.4: Poems comprising *ARCorp* …… 117
Table 3.5: Poems comprising *MVCorp* …… 118
Table 3.9: KW data comparison between *PPCorp* and *MEPoetLitCorp* 132
Table 3.11: Subjectivity- and limitation-based challenges to annotation 135
Table 3.12: Amounts of post-annotated modernisation …… 137
Table 3.15: Approach to data analysis …… 144
Table 3.16: Top 100 most frequent words in *PPCorp* and *MEPoetLitCorp* 146
Table 3.17: Top 50 lexical KWs generated for *PPCorp* …… 150
Table 4.1: Top 20 most frequent words in *PPCorp*, *MEPoetLitCorp*, and *EEBO* …… 160
Table 4.2: Top 50 most frequent lexical words in my corpora …… 165
Table 4.4: Primary thematic KW domains in *PPCorp* …… 171
Table 5.1: Consumption-based keywords …… 179
Table 5.3: Collocates for the KW *hunger* in *PPCorp* …… 189
Table 5.4: Collocates for the KW *glutton* in *PPCorp* …… 190
Table 5.9: Collocational animacy with *hunger* …… 201
Table 5.11: Collocates for the KW *drink* in *PPCorp* …… 211
Table 5.15: Types of evaluation …… 225
Table 5.17: Evaluation of *drink* within *PPCorp* …… 213
Table 5.18: Social esteem and social sanction …… 235
Table 6.1: Legal keywords …… 256
Table 6.4: Collocates for the KW *justice* in *PPCorp* …… 268
Table 6.5: Collocates for the KW *doctor* in *PPCorp* …… 269
Table 6.10: Paxson's (1994) figural translation taxonomy …… 282
Table 6.12: Collocates for the KW *reason* in *PPCorp* …… 288
Table 6.14: Classification of the *PPCorp* legal keywords by technicality 297
Table 6.16: Legal binomials in *PPCorp* …… 309

List of appendices and other content

Appendix A: Information on the source-texts for each corpus

CD: Listing of the 855 keywords generated for *PPCorp*

1. INTRODUCTION

1.1 Why study *Piers Plowman*?

Piers Plowman occupies a unique position in discussions of late medieval English literature and language. Attributed to William Langland, this otherwise anonymous 14th century poem exists alongside Chaucer's *Canterbury Tales,* Gower's *Confessio Amantis,* and the *Pearl* poet's *Sir Gawain and the Green Knight* as the cornerstones of late Middle English poetic literature. This pre-eminent literary status is evident in – and no doubt perpetuated by – the poem's established permanence in undergraduate curriculums, and further supported by the substantive amount of critical literary study surrounding the text.

What makes *Piers Plowman* compelling as a literary text is its linguistic creativity. Described as 'deploy[ing] a language of unusual richness, density, and allusiveness' (Pearsall 2006), the poem exhibits a characteristic textual complexity which underlies its reputation as being a notoriously difficult text to analyse. Literary critics repeatedly comment on the linguistic atypicality of the poem in relation to other contemporary works, with 'all of the alliterative poets except Langland [having written] within a single remarkably stable poetic' (Duggan 1986: 569). Studies in both literary criticism and historical linguistics have critically examined the poem's textual features with particular focus on grammar (Samuels 1985, 1988; Horobin 2005), metre (Duggan 2009, Moriya 2014, Weiskott 2016), and lexis (Oakden 1935, Turville-Petre 2013).

This research seeks to study the language of *Piers Plowman* using computational approaches developed in corpus linguistics. Part of a wider movement observed in the digital humanities (Jensen 2014: 130), a corpus approach offers new ways of examining the linguistic elements contributing to the 'textual knottiness'

(Pearsall 2006) associated with the poem's style by computational identification of recurring textual features compared to the language evidenced in a collection of 28 other late Middle English poetic texts. Based on statistical comparison of language use in both sets of data, this approach facilitates linguistic examination on an empirical basis and has the potential to reveal patterns both a) shared with other contemporary literary works, and b) distinctive to (and within) *Piers Plowman.* The textual (dis)similarity evidenced in the corpus data provides a starting point for analysis regarding the poem's literary style, which constitutes the focus of the present investigation.

From among the stylistic patterns evidenced in the text of the poem, this research focuses on the language patterns embedded in the representation of thematic consumption and the law. This approach acknowledges the demonstrable interest in these topics found in Langlandian scholarship (e.g. Mann 1979, Schmidt 1980, Savage 1993, Benson 2004, Kim 2004, Gasse 2016 for consumption; Alford 1977/1988a, Stokes 1984, Baldwin 1981a/1981b, for law and legality); it also takes into account the evidenced amount of statistical significance attributed to the lexis used to convey these themes in the text, as revealed through the corpus data generated in this study.

Exploring the poem's language through corpus methods is intended to complement the findings of existing linguistic and literary study of the poem, with the use of corpora orienting analysis around the text's lexis. While unlikely to reveal key terminology not already identified by Langlandian scholarship, a corpus approach provides the means to examine the underlying lexico-grammatical patterns within which this language is contained and offer insight into how these are furthermore used for stylistic effect. These patterns are accessible through approaches such as keyword

analysis, which measures the unusual frequency of occurrence of a word within a given dataset (Culpeper & Demmen 2015) and collocational analysis, which assesses the strength of affinity between words (or sets of words) based on the likelihood of their textual proximity (Xiao 2015). These statistically-based measures provide access to a level of insight into the text's recurring language patterns that would be difficult to achieve through manual analysis alone (Carter 2010: 65; 2012).

My approach thus situates this investigation within the field of corpus stylistics (Mahlberg 2014, 2016). Corpus stylistics combines quantitative and qualitative analysis to study style as literary effect, whether perpetuated across a literary work (Stubbs 2005, Mahlberg 2010, Ho 2011, Mastropierro 2017) or expressed in sections of a given text (eg. Culpeper 2002, 2009). Underlying this approach is the conviction that applying corpus linguistic techniques to the study of literary texts provides insight into the relationship connecting the linguistic patterns present in the text with the literary impact and appreciation experienced by its readers (Mahlberg 2013: 2).

Performing a corpus stylistic investigation of *Piers Plowman* is groundbreaking in that it departs from traditional Langland studies, as well as current research in corpus stylistics. The poem's presence in multiple surviving manuscripts has directed scholarly attention towards manuscript histories and textual comparisons of different versions of the poem (Pearsall 1990: 6-32), with a comparatively smaller sphere of study dedicated to examining the text's language-based literary style. Similarly, corpus stylistic approaches remain rare in the investigation of Middle English literature; despite the overwhelming disciplinary orientation of literary stylistics on modern literature, recent corpus-based studies have featured analysis of Shakespeare's language (e.g. Busse 2007, Culpeper 2009, Anderson & Crossley 2011, Demmen 2012, Murphy 2015) but have yet to tap into the wealth of linguistic information available in

the late medieval texts composed nearly two hundred years earlier. My research addresses both of these under-represented areas of scholarly inquiry, and has the potential to contribute to existing research in both spheres.

1.2 Research questions

The overarching research question (RQ) guiding the investigation asks: ***how do the lexical features revealed in a corpus analysis of Piers Plowman contribute to analysis of the poem's literary style?*** This focus can be broken down into four sub-questions, as outlined below in (i-iv):

(i) *What are the lexical similarities and differences evidenced between Piers Plowman and other late medieval poetic texts, as revealed by the corpus data?*

(ii) *From amongst the lexical features deemed distinctive of (or unique to) Piers Plowman, which terms address the themes of consumption and the law in a late 14th/early 15th century context?*

(iii) *Within this legal and consumption-based lexis, what lexico-grammatical patterns emerge and how do they contribute to observable stylistic effects in the poem?*

(iv) *How do the insights provided by these stylistic patterns meaningfully contribute to (or corroborate) discussion of legal and consumption-based themes in the existing literary criticism?*

The first two questions (i and ii) are answered by comparing the poem's language to a reference corpus of Middle English literary text, while questions (iii) and (iv) are

predominantly addressed by analysing the language features and patterns within *Piers Plowman*. Examples are occasionally drawn from the reference corpus to answer (iii) and (iv), however their use is predominantly illustrative to provide comparative perspectives to supplement close analysis of the poem. The approach taken in answering these sub-questions (i-iii in particular) is fundamentally data-driven (eg. Rayson 2008), with the selection of textual features for further examination based on the initially-generated corpus results.

This investigation additionally engages with methodological considerations concerning the historicalness of the literature (and by extension, language) being analysed. In this capacity the project is guided by a secondary research question: ***how does the historical nature of the late 14th/early 15th century poetic texts being used impact a corpus approach to language analysis?*** Like the main RQ, this can be further divided into the three sub-questions set out in (v-vii) below:

(v) *What are the challenges associated with analysing Middle English using corpora, and how do these affect the use of modern corpus tools and software?*

(vi) *How should these issues be addressed to generate viable results for keyword and collocational analysis in Middle English texts?*

(vii) *How effective are these methods at providing usable corpus material, and how can these results be further improved?*

While not the primary focus of this project, this supplemental RQ examines the obstacles facing the application of corpus linguistic approaches, tools, and theoretical frameworks to Middle English (v), strategies for improving the accuracy of the generated corpus results (vi), and a critical appraisal of the degrees of effectiveness

observable in my own study (vii). The insights garnered from this line of investigation are meant to contribute to a growing body of scholarship on digital approaches to historical varieties of English (e.g. Markus 2000, Baron & Rayson 2009, Lehto et al. 2010, Archer et al. 2015, Jenset & McGillivray 2017).

Taken collectively, these research questions set out an approach to the poem's language that is analytically innovative and methodologically reflective. RQ1 (as set out in i-iv) explores how recourse to corpus-based theory and methods facilitates the study of linguistically-conveyed style in the text, while RQ2 (as outlined in v-vii) appraises the techniques by which to best address the challenges associated with taking a corpus linguistic approach to Middle English language.

1.3 Thesis structure

This thesis is comprised of seven chapters, including the present section. Given the inter-disciplinary nature of the project, the following chapter (Chapter 2) deals extensively with the range of theoretical perspectives harnessed in its investigation. It locates the present study within existing scholarship by highlighting the three areas informing my research: literary linguistics, corpus linguistics, and the historical and literary critical tradition surrounding *Piers Plowman* studies. In showing their interconnectedness I argue that all three are needed for an analysis of the poem that is contextually informed, grounded in relevant theory, and engaging with topics of interest in current Langlandian scholarship. This chapter also contains some background information on the poem itself, as well as a brief outline of the literary context surrounding its composition.

Chapter 3 outlines the mixed-methods quantitative and qualitative approach taken in this investigation in light of each discipline discussed in Chapter 2. From a corpus linguistic perspective I discuss the decisions made in the construction and annotation of the *Piers Plowman* target corpus (*PPCorp*) and its poetic reference corpus (*MEPoetLitCorp*). I address the practical implications of applying corpus methods to literature written in Middle English, and use select examples from the annotated corpora to signpost some of the factors contributing to the resulting quality of the generated corpus results. This chapter also acknowledges the literary and textual considerations involved in the selection of the version (and, by extension, source-text) of the poem used in the present investigation and the potential implications resulting from these decisions. I additionally justify the corpus-based focus on keywords (and the collocational patterns in which these keywords are situated) from a literary linguistic perspective by signposting their function as markers of literary style (Mahlberg & McIntyre 2011: 207).

Chapter 4 provides a quantitative survey of the lexical landscape of *Piers Plowman* compared to the medieval literary reference corpus. Using the frequency and keyword data generated in the comparison of these corpora, this section highlights some of the lexical features characteristic of these Middle English narrative poems; it also discusses some of the fundamental differences between poetic types (i.e. alliterative versus metrical), based on distinctions rendered visible in the corpus data. The section then concentrates on the top keyword results for *Piers Plowman* to narrow analytical focus for the remainder of the study to a) two of the more salient thematic keyword categories, and b) the theoretical linguistic frameworks best suited for analysing each selected keyword category.

Chapter 5 contains the first of two analysis chapters, with its content focused on the keywords containing the poem's consumption-based lexis. Within the generated set of keywords it examines the language of dietary consumption in *Piers Plowman* as framed within the context of evaluative language use (Martin & White 2005), drawing attention to the stylistic significance of the collocational patterns evidenced in the concordance lines in which these keywords appear. In addition to stylistic observations, the chapter reflects on the durability of the literary linguistic theories and frameworks being used, with suggestions made for adaptation where applicable. This corpus-based stylistic analysis is then used to link the present study to ongoing discussions regarding consumption within the poem's existing literary crticism, with exploration focused on the way(s) by which the corpus findings corroborate, build on, or otherwise contribute to existing study of the text.

Chapter 6 continues this analysis of the poem's language by shifting focus to keywords relating to legality and the law along a cline of linguistic technicality (Marin & Rea 2014, Marin 2014). As with the consumption-based keywords, I first conduct a stylistic analysis of the keywords and their corresponding collocational patterns before integrating my findings with literary scholarship examining legal terminology and language in *Piers Plowman* (e.g. Alford 1977, 1988a, Rudd 1994, Galloway 2001).

The final chapter of this thesis (Chapter 7) provides a general summation of the project's main findings and outcomes alongside discussion of limitations encountered in the current study and possible avenues of future research.

2. LINGUISTIC APPROACHES TO HISTORICAL STYLE: EXAMINING *PIERS PLOWMAN* THROUGH CORPUS ANALYSIS

2.1 provides a brief introduction to *Piers Plowman*, orienting it within the literary and historical contexts of its late fourteenth century composition as well as highlighting broader trends characterising approaches taken to the text in Langlandian scholarship from the late nineteenth century onwards. 2.2 establishes the investigation's theoretical parameters by situating my examination of the poem's language within a (historical) literary linguistic framework. This section examines the various conceptualisations of style evidenced in both modern applied linguistic and critical literary research and highlights their distinctiveness from the approach to style currently being utilised. 2.3 examines the role of this project relative to existing research in historical corpus stylistics, and demonstrates how access to corpus theory and methods (specifically, keyword and collocational analysis) assist in identifying the linguistic features on which my stylistic analysis of *PP* is based. 2.4 concludes the chapter by considering some of the additional challenges facing a corpus analysis of Middle English text.

2.1: Piers Plowman

A narrative poem comprised of over 70,000 words of unrhymed alliterative Middle English verse, *Piers Plowman* (hereafter, *PP*) constitutes a dense piece of poetic literature the size of a short Modern English novel. Thought to have been composed in the late 14th century (approx. 1370-1390), the poem's attributed authorship to William Langland is based on evidence from a single 15th century manuscript (Trinity College MS 212) ascribing the text to a 'William de Langlond', as well as possible biographical

information inferred from the author's professional and experiential knowledge evidenced in the content of (and language used in) the text.[1]

PP features within a strong tradition of early English literature predating the English Renaissance. Operating within a period of 'poetic florescence' (Burrow 1971: 1), the literary landscape in the 14th and early 15th centuries was dominated by narrative poetry in a tradition which, according to Burrow (2008: 60), was characteristic of 14th century literature in the same way the novel defined 19th century English literature. From within this poetic tradition emerges the earliest group of stylistically 'recognisable' English literary writers, including Chaucer, Gower, and Langland (Burrow 2008: 46).

This late medieval literary movement features two separate poetic traditions, metrical and alliterative poetry. The metrical poems, as exemplified in the works of Chaucer and Gower, are characterized by strict adhesion to set metrical patterns, while the alliterative tradition (to which *PP* belongs) is comprised of unrhymed lines of no set length but containing alliterative schemes wherein consonant-based sounds appearing at the beginning of a word or stressed syllable are repeated within the same line for stylistic effect (Short 1996: 107-11). Texts like *PP* belong to the Alliterative Revival (Turville-Petre 1977, Chism 2002), a wider literary movement including texts such as the alliterative *Morte Arthure, Mum and the Sothsegger, Wynnere and Wastoure, Sir Gawain and the Green Knight*, and *St. Erkenwald*, and is thought to be connected to the earlier alliterative traditions evidenced in Old English poetry.[2]

[1] For an example, see Schmidt's (1987) discussion of Langland as a 'clerkly maker', a title evidenced in the combination of legal and poetic knowledge reflected in the poem.

[2] There is some debate regarding the historical relatedness of OE and ME alliterative poetic traditions. Spearing (1970: lxvii), reiterating Chambers (1932), argues that the alliterative poetry seen in Middle English represents an unbroken tradition of alliterative verse stemming back to Old English through a continuous oral tradition. However, sources including Turville-Petre (1977) claim that the lack of

In addition to its metrical tradition, individual features ascribed to *PP* situate the text within a series of overlapping medieval literary genres. For example, its use of personified allegorical characters (in addition to broader types of allegorical projection) aligns the text with other poetic allegories (for example, *Romance of the Rose*). Similarly, the embedded narrative structuring situates *PP* alongside other dream visions (*Romance of the Rose, Pearl*) while its extended dialogic content between opposing interlocutors operates as a characteristic element of debate poetry (*The Owl and the Nightingale, Parliament of Fowls*). The poem moreover exists as the attributed precursor to the *Piers Plowman* tradition, a literary movement containing poetry composed between 1393 and 1414 reminiscent of *PP* in the recall of key themes, phrases, and referenced episodes in texts such as *Pierce the Plowman's Crede, Richard the Redeless, Mum and the Sothsegger*, and *The Crowned King* (Barr 1993: 6-7).

2.1.1 The text

The plot of the poem follows a first-person narrator named Will through a series of embedded episodic dreaming and waking periods. During his dream-visions Will encounters an array of allegorical personifications featuring virtues such as *Truth* and *Reason*, vices including *Sloth* and *Gluttony*, and institutions such as *Holy Church* (i.e., the Christian faith). The action of the poem is predominantly based on the dialogue between Will and these allegorical characters or his passive observation of the interaction of said characters amongst themselves, with the focalisation on a first-person narrator inviting the reader – through the dreaming narrator – to experience

---

documented evidence makes assertions of continuation over reinvention difficult to either refute or substantiate with any great confidence. See Weiskott (2017) for updated discussion.

the discoveries seen, urgencies shared, events witnessed, and revelations experienced over the course of the poem (Pearsall 1994: 17).

The longest version of the poem contains a prologue followed by twenty-two passus, or poetic 'chapters'. The prologue starts with the dreaming narrator finding himself in a field surrounded by a group of representative social and professional figures enacting their role(s) as self-serving individuals. He meets Holy Church in passus 1, who explains the vision to him and engages Will in philosophic debate on acceptable behaviours under religious doctrine and ways in which the soul can be saved through charitable deeds and action. Passus 2-5 contains an allegorical portrayal of the corrupting influence of money on society with the introduction of the character Lady Mede and her attempted influence over all with whom she comes in direct contact. Passus 6-7 follow up this allegorical vision with a separate allegorical scene wherein the Seven Deadly Sins confess their crimes to Repentance, who prays for their general forgiveness. This acts as the springboard for the collective search for Truth in passus 8, where the dreamer encounters the figure of Piers the plowman. Passus 8 and 9 continue to follow Piers and his crowd of labourers as they plant crops to create their well-organised and self-sufficient society, all the while engaging in overarching discussion of the promises of salvation through honest work and the need for inward and self-generated reformation for those inclined to avoid work instead of the external coercion manifested through forces such as Hunger and Poverty.

From passus 10 onwards, the poem shifts into a series of temporal experiences and stories merged together under the same ideological drive towards achieving (and understanding the underlying processes of) salvation. The dreamer goes on a search for Dobet (or 'do-well') within himself in passus 10 through interactions with Thought and Wit as elements of his own intellectual faculties, who provide answers to

questions of salvation through good actions and behaviour supplemented by suggestions made by allegorical figures such as Study, Clergy, and Scripture in passus 11. Passus 12 sees the narrator, overwhelmed by the amount of information presented to him, have an episodic dream wherein he had fallen into the clutches of Recklessness and squandered his life away in worldy service to Fortune and Covetouousness, a situational predicament enabling further discourse on the relationships between financial security, gainful employment, and the exercising of conscience in the lives of the individual as well as on a social level more generally. These concepts are then applied to the means by which to achieve spiritual harmony in passus 13-14, predominantly through enaction of humility and charity in accordance to religious doctrine. Passus 15 seees the narrator attending a feast and acquiring Patience as a mentor, with passus 16 exploring the ideology behind poverty as a method of attaining grace and understanding truth with the assistance of the Church. Passus 17-19 contain discussion of a vision of grace for mankind heavily supplemented with references to Biblical stories and figures which, through the merging of the ideal Good Samaritan persona with the allegorical figure Charity and the fictional character Piers, introduces a veritable Christ figure as the embodiment of salvation.

The final three passus of the poem focus on revelation. Passus 20 contains the arguable climax of the poem through accounts of Biblical stories including Christ's crucifixtion and the harrowing of Hell. The subsequent establishment of the Church and its function as the machinery of redemption is expounded on at length in passus 21, with the narrator extolled to maintain his search for truth in passus 22 through the image of the Church and its personified defenders being attacked by oncoming waves of allegorical sins and lesser temptations. The poem ultimately ends with Conscience vowing to become a pilgrim and wander the earth seeking out Piers the plowman, the

final memory retained by Will before reawakening from his dream to his real-world surroundings in the Malvern Hills.

2.1.2 Features of the text

Through its presentation as a dream vision *PP* provides a satirised representation of mankind's struggle for salvation, one that operates as a highly-stylised version of the conflict visible in the historical context of its composition. Fourteenth (and early fifteenth) century England represented a tumultuous period characterised by political, religious, and social communal tension and upheaval (Du Boulay 1991: 5-7, Barr 1993: 1-8). The uncertainty surrounding Richard II's underage accession, misrule, and subsequent deposition in favour of Henry IV and later Henry V appeared concurrent with trepidation over military fears caused by ongoing civil disturbances, open Welsh rebellion, and the threat of French invasion; these military problems were exacerbated on a socio-economic scale by the amount of taxes levied to fund the secular government. Such events occurring on a national level paralleled ongoing conflict targeting specific institutions like the issues faced by the Christian church brought by the followers of John Wyclif who, among other taken theological stances, highlighted the corruption of the institution and actively advocated for ecclesiastical disendowment. Operating beneath the surface of these larger activities were demographic changes in a post-plauge society characterised by greater urbanisation and literacy, which facilitated movement away from earlier feudal economies and provided opportunities for greater social mobility. From amongst these new professional roles as merchants, lawyers, and small landowners the Lollard movement arose, which was characterised by greater communal awareness of (and engagement

with) political events and new ideas of religious practice – such as access to the Bible in English in defiance of existing educational and salvational processes – than observed previously.

The poem's active engagement with contemporary events and shifting fluctuations of power and authority in 14th century England is well-documented in Langlandian scholarship. Areas of particular interest have concentrated on the text's portrayal of ecclesiastical politics (Aers 1975, 2015; Adams 1988), the state of legal (Alford 1977, 1988a) and governmental (Baldwin 1981) affairs, and other socio-cultural issues.[3] While this relationship between text and context can be generally characterised in *PP* as the means through which the narrator achieves higher spiritual understanding via an investigation of contemporary society, Barr (1993: 4) also identifies the role of the text as a physical artefact embedded in ongoing real-world events as a source that 'not only commented on matters or church and state, but was itself caught up in them…as a watchword for civil disobedience'. Pearsall (1994: 13) corroborates this role by observing how the poem records – and in some instances, directly critiques – the institutional change that was itself ongoing during the time of *PP*'s composition:

> The poem records in the minutest detail the conflict which racked late medieval society, as the feudal order and the Church of the west moved into their last stages of institutionalised decay, and as the antagonistic forces over which they had presided moved into the open area. The strains and pressures between the shifting social strata of society – between government and people, lords and commoners, clergy and laity, possessioners and mendicants – had built up during a long period of apparent stability, but the later fourteenth century saw the first release of tension, the first open fracture in the fabric of the traditional order, in the two events which Langland's poem provides both preclude and commentary – the Peasant's Revolt of 1381 and the condemnation of Wyclif's teaching in 1382.

[3] For instance, Crassons' (2010) exploration of the ways in which the poem portrays the relationship between poverty and medieval communal identity.

Because of its prominent historical contextualisation the poem is often studied alongside contemporary political discourses (i.e., Robbins 1960, Aers 1980), with the argument being made that the highly politicised nature of the poem was meant to be read and interpreted in light of its surrounding ideological context. Scholars including Barr (2014) rightfully note that, in many instances, the text addresses more than one of these contextual factors (sociocultural, political, legal, economic, religious) simultaneously, a situation generating a degree of thematic complexity that has come to characterise the text within its critical scholarship.

Alongside the acknowledged complexity of the poem's thematic content, the poem also contains distinctive linguistic patterns and literary structuring compared to other contemporary poetic works. Salter and Pearsall (1969: 2) identify the poem's absence of commitment to a set or otherwise continuous narrative structure as being unusual, while Kirk and Anderson (1990: ix-x) note that the text uses lengthier alliterative lines than conventionally seen in Middle English alliterative poetry. It has also been observed that the poem makes recurring use of atypical intra-linear alliterative pairs between content and function words, rather than placing alliteration solely on stressed lexical terms as traditionally done in both Old English and other Middle English poems (Robertson & Shepherd 2006: xi).

Another feature characterising *PP* is the large volume of Latin distributed throughout the otherwise English poem. The presence of Latin in *PP* is to some degree reflective of the diglossic linguistic context of the poem's composition in a post-Norman Conquest England, with Latin, Middle English, and Norman-French each used for different functions within separate professional and situational contexts (Sylvester 2015: 589-90). Appearing in both narration and within character dialogue, these Latin

passages typically operate as quotation-based instances of code-switching involving the substitution of single words, isolated phrases, or even entire sentences at a time:

(i) Thenne scripture scorned me and mony skiles chewed
…And saide **multi multa sciunt & seipsos nessiunt** (11:171,174)[4]

Then Scripture scorned me and made many arguments
*….And said **'many man know many things but do not know themselves'**.*

This linguistic duality is not sporadic in *PP* but rather represents a definitive textual feature throughout the poem, as well as within late medieval written discourse more generally (Schendl 2000). While several other late medieval poems contain occasional lexical substitutions or phrases in Latin and French (for example, Chaucer's *Canterbury Tales*), the amount of non-English text contained in the poem is only overshadowed by Gower's *Confessio Amantis*, which makes liberal use of both French and Latin simultaneously.

2.1.3 Versions of the text

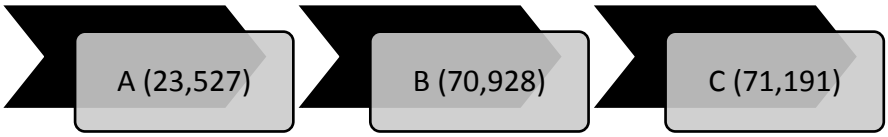
PP possesses a complex manuscript history based on its presence in multiple surviving copies, each containing versions of the poem differing from one another in both content and overall length. The poem currently survives in approximately 53 manuscripts appearing with varying degrees of completeness and textual (dis)similarity to one another.

To provide a classification system for these manuscripts I follow the majority of modern Langlandian scholars in adopting Skeat's (1888) work outlining three 'distinctly authorial' forms of the text: A-text, the B-text, and the C-text. Their gradually

[4] All textual extracts reproduced in this thesis are drawn from the *X* manuscript for *PP* (see 3.1.2).

increasing lengths, in addition to their textual similarity and signs of progressive revision, constitute the basis for the ABC cycle, a widely-held theory that these three manuscript traditions collectively represent a chronological progression of authorial revision over approximately twenty years. These versions are believed to be cumulatively representative of the author's influence and overall direction (Skeat 1888: ix), as shown below in Figure 2.1.[5]

Figure 2.1: Chronological representation of the ABC cycle of the *PP* manuscript tradition with approximate word counts.



These identified authorial versions of the poem get progressively longer based on several observable differences within the text. The A-text contains a dozen passus, work which was extensively rewritten then further tripled in length with the addition of nine more passus in the B-text. The C-text constitutes a more radical and sporadic revision of the B-text, with significant additions and omissions across the poem (Pearsall 1994: 9-10). In these instances, the use of terms such as 'C-text' refer not to a specific manuscript but rather to a tradition of manuscripts that medieval textual scholars and book historians have determined possess greater similarities than differences to one another according to the frequency of perceived coincident or convergent textual variation. These types of proximal relationships, or 'genetic relations' (Russell & Kane 1997: 19-20), are dependent on (and provide an illustration

[5] A few scholars have argued against this cycle (i.e., Mann 1994), with some advocating for the inclusion of an additional authorial manuscript tradition – the Z-text – within the existing ABC cycle (Rigg & Brewer 1983, Brewer 1996). However, these perspectives do not possess as staunchly robust followings in the literary criticism and are therefore not addressed here.

of) which version was copied from which earlier manuscript, and at what point(s) in time.

2.1.4 Modern reception

Modern readers typically encounter *PP* within a higher educational context, either as fragments of written material through which students learn Middle English or holistically as a piece of poetic literature in its own right. However, despite being a foundational pillar of late medieval English poetry, the poem does not tend to be read with great frequency beyond the academic sphere. Reasons for this could be attested to modern readers' cognitive partiality for prose over poetry in lengthy narratives (i.e., novels over epic poems), cognitive gaps in experiential schema separating modern readers from the poem's 14th context of composition, and a modern dispreference for stylistic features such as alliteration or allegory in contemporary poetic traditions. Broadly affecting Middle English poetry, these preferences appear to be surmountable through written and cinematic adaptations targeted at popular culture audiences as done for texts like *Sir Gawain and the Green Knight* and *The Canterbury Tales*, but comparatively few of these presently exist for *PP*.

In addition to modern knowledge gaps and literary preferences, another factor contributing to the poem's reception is the perceived absence of structural cohesion and semantic ambiguity within (and across) the poem. The absence of cohesion in the text's grammar, use of proper nouns, and its narrative structure results in issues relating to coherence, or lack of interpretive clarity. These problems with overall coherency are visible in the brief summation of *PP* provided in 2.1.1, which reveals how the key events taking place in the poem are scattered across nested dream visions

and erratically jump between levels of fictional social and self-introspection with minimal signposting. This cohesion-facilitated absence of ideological coherency can be linked to the combined effect of the poem's linguistic and and thematic complexity, and is no doubt exacerbated by the at times dramatic structural changes evidenced within the plurality of available versions of the text.

The poem's coherency (or arguably, lack thereof) has been recurringly identified in its surrounding literary criticism as a characteristic feature of the text, albeit one initially viewed with critical reception. Pearsall (1977) summarises the poem by evaluating its compositional elements with a notable degree of descriptive baldness:

> …the structure of the poem is associative and idiosyncractic, the very sequence of materials often difficult to understand, its handling of dream and allegory shifting, inconsistent opportunistic; what appears to be its main ordering structure, [the main character, Will's, search for]…Dowel, Dobet, and Dobest, turns out to be a façade, and the central theophanic character, Piers Plowman, a mystery; characters emerge, disappear, re-emerge, problems are taken up and dropped unsolved. By any standards but its own it is near to artistic breakdown. (Pearsall 1977: 178, cited in Robertson & Shepherd 2006)

This unreservedness in the critical reception of the poem echoes Muscatine (1963: 121) who, in addition to observing the 'extraordinary inconclusiveness' in the text, has more contentiously described the poem as having 'a curiously homogeneous artistic effect that for a lack of a better term I have called surrealistic'.[6] This description of the effect produced by reading the poem, Muscatine (1972: 106, as cited in Davlin 2018: 115) argues, is primarily attributed to the presence of abrupt transitions, semantic juxtapositions, and philosophical meanderings.

[6] See also Alford (1988c: 29-32).

However, more recent scholarly approaches to the text have argued for a reconciliatory middle ground replacing the poem's traditionally-maintained reputation as 'illogical, unstable, and enigmatic' with an approach 'that recognises its artistic power and achievement at the same time it acknowledges the demands its textual history and narrative texture impose' (Economou 1996: xviii). This more moderate approach to the text prioritising contextual-based readings over the text as an isolated artefact has since become the default position of current critical research, as outlined below in the following section.

2.1.5 Critical scholarship

Critical scholarly approaches to *PP* have developed along a chronological trajectory in the amount of interest paid to the text, as well as the type of research conducted. Economou (1996: xvi) outlines some of the broad trends characterising the historical progression of Langland studies to date, starting in the early 19th century with the poem featuring as 'a major document in England's social history, a record of complaint and satire against the abuses of church and state, a proleptic symbol of the triumph of democracy and the common Christian man' (see also Simpson 2007: 1). This status, Economou (1996) observes, changed in the latter half of the century with Skeat's (1888) classification of the poem's surviving versions into set manuscript traditions. In the process of reinvigorating studies of the poem in modern scholarship, Skeat's (1888) work also guided early 20th century scholarship into authorship attribution studies, with the field's general consensus on William Langland as attributed author largely established by the 1950s by consensus rather than for purposes of referential convenience (Kane 1965). This cumulated in latter 20th century scholarship focusing on

'questions pertaining to style, structure, unity, and coherence of the poem' (Economou 1996: xvi) which has manifested in research involving comparative literary study (how (dis)similar is the poem to other medieval dream visions or allegories, for instance), biblical exegesis, social and intellectual histories addressing the text's late 14th century context (i.e., a historiographic approach), and manuscript studies.[7]

These latter areas of scholarship – historiography and manuscript studies – represent the most prolific branches of modern *PP* research. Each area features strong traditions of critical work dedicated to either orienting the poem within the historical context of its composition, or otherwise exploring the textual differences between versions of the poem within (and across) each established manuscript tradition. Historiography studies, referred to by Middleton (1988: 19) as the 'social contexts of literary practice', traditionally constitute topical examinations of thematic ideology connecting the poem to its late 14th century discourse world exemplified by studies such as Baldwin (1981) and Crassons (2010). An example of a manuscript studies, or book history, approach can be found in the critical discussions embedded in the *Athlone* editions of *PP* (Kane 1960, Kane & Donaldson 1988, and Russell & Kane 1997) which, in addition to providing modern edited versions of the text, also explore in great detail the comparative textual differences between manuscriptal relationships evidenced within each tradition of the poem (A, B, and C, respectively).

The critical approaches applied to *PP* have been diverse across the poem's historical trajectory but the most prominent are the political readings brought by Aers (1980), Du Boulay's (1991) historical approach, and the theology-based ideological reading introduced by Simpson (2007). This triumvirate functions within Langlandian

[7] For a more comprehensive overview of the historical critical heritage surrounding *PP* (i.e., pre-late 19th century views and readings of the text), see Middleton (1988).

scholarship as essential readings for further study of the poem, texts which are being supplemented by the currently-ongoing development of the *Penn Commentary on Piers Plowman* which, through its envisioned five volume series, is meant to offer the first updated commentary of the poem published since the nineteenth century and is intended to operate as a key resource for Langland scholars upon its completion.[8] My current study of the poem is situated within (and contextually informed) by the information provided in these these sources to ensure that my investigation addresses general trends, findings, and interpretations within the poem's mainstream literary criticism.

Within this critical remit, my stylistic approach to the poem is fundamentally different to previous language-oriented studies of the poem. Historical linguistic approaches to *PP* traditionally tend to be theoretical in their focus on philology, with interest predominantly based on features such as poetic metre (Duggan 2009, Moriya 2014, Weiskott 2016), lexis (Rudd 1994, Turville-Petre 2013), and dialect studies (Samuels 1985, Duggan 1990). The present investigation departs from this tradition of linguistic examination in taking an applied linguistic approach to the text using corpus linguistic tools and methods to explore literary style as represented through the poem's lexico-grammatical patterns. This does not mean that 'style' has not been addressed in Langlandian scholarship up until this point but rather that my linguistic approach is situated outside the theoretical scope and tradition characteristic of *PP* studies, and approaches to historical language more generally (for more discussion, see 2.2).

[8] The first three volumes of the Penn Commentary have already been published upon the date of writing; these works cover the poem in sections, with Galloway (2006) addressing content and scholarship from the prologue to passus 4 (volume 1), Hanna (2017) covering passus 5-9 (volume 2), and Barney (2006) containing discussion relevant to passus 18-22 (volume 5).

Despite the more recent changes in scholarly focus identified by Economou (1996), study of the poem's style has been overshadowed by the overwhelming dominance of manuscript studies in linguistic examinations of the text. Both areas examine the poem's language but with different scholarly objectives: manuscript studies focus on the comparative differences in language use for the purpose of identifying holographs (manuscripts written by the attributed author themselves) from prototypes (the most complete, 'clean', or otherwise closest surviving texts to the original holograph) and corruptions (texts derived from prototypes to varying degrees of proximity) containing 'errors' (i.e., document changes) not visible in the original text. By contrast, stylistic inquiry is less concerned about the relative purity of the text as defined by its proximal historical relationship to the attributed author and its other manuscriptal versions than on the linguistic techniques and patterns occurring within the poem itself.

Investigations of style in the critical scholarship predominantly focus on the text's poetic form, or the linguistic structuring of the poem. Lawlor (1962) offers one of the earliest modern explorations of poetic form through systematic consideration of the text's metrical form, rhythmic patterns, allegory, and wordplay, with the latter signalling an interest in the functional deployment of lexis to generate new meaning. Schmidt (1987) continues this survey of the poem's style through discussion of the 'poetic arts', which echoes Lawlor's (1962) interests in looking at Langland's versecraft (metre and rhythm), diction (style-based selection of lexis), and wordplay. Further evidence of style-based research in *PP* scholarship, even if not explicitly titled as such, can be found in more function-oriented studies examining stance (Barnes 1995), allegorical agency (Donner 1987), and poetic texture and tone (Muscatine 1972: 93-4).

My use of computational methods to further examine the language-based style evidenced in the poem aligns with the current directionality of the field towards digital resources. This does not presuppose a previous absence of computational tools and resources in facilitating literary analysis (Hockey 2000: 67, cited by Mahlberg 2009: 48) but rather indicates paralleled innovative development wherein ongoing technological advances in language analysis are proving useful to answering new questions in literary exploration. The emerging proximity of current Langlandian studies to corpus research is visible in both the computationally-assisted production of a print concordance (Wittig 2001) as well as the ongoing development of the *Piers Plowman Electronic Archive*, a volunteer-led academic initiative to produce a publicly-available annotated corpus containing copies of the different manuscript versions of the poem on a single digital platform for comparative study.

This broader movement towards the digital, in addition to the comparative absence of literary stylistic analysis in *PP* scholarship generally, suggests the suitability of the present investigation as not only being a project that collaboratively engages with new approaches gaining traction in the field but also one that offers new insights to a rich tradition of existing critical literary scholarship. The developing interface between literary analysis and digital methods in Langlandian studies is particularly useful in systematically examining the poem's themes and how they are constructed linguistically. A benefit to this approach is that it enables further work to be conducted on themes already identified as being significant within the critical literature surrounding the text, with themes of legality and consumption being two areas of critical focus discussed in existing Langlandian scholarship.

2.1.6 Themes in Piers Plowman: consumption and the law

While not the only themes addressed in *PP*, consumption and the law represent two areas that have generated substantive interest in the text's critical scholarship. Interest in consumption is based on the topicality of the subject to the poem, with numerious characters, events, and philosophic dialogue directly addressing aspects of consumption through the existential conditions surrounding under- or over-consumption (poverty and excess, respectively). By contrast, the focus on legality in *PP* lies in the high occurrence of specialised legal terminology found in the poem (Alford 1988a), a presence which has been used to inferentially bestow the author with extensive knowledge of the 14th century courts and English legal system.

The following sections set out each of these themes in greater detail by contextualising their significance from a 14th/15th century historical perspective, as well as through the relevant scholarly work conducted by Langlandian scholars and other medievalists.

2.1.6.1 Consumption

Consumption has recurringly been a subject of interest within a late medieval context, particularly for establishing patterns of consumerism characterising conventional social practices. Studies focusing on late medieval periods of English history examine evidence of dietary consumption from surviving household account ledgers and diaries to determine caloric intakes typically consumed by the affluent (Woolgar 1995, 2001) versus the general populace (Dyer 1983, 1998), while others look at consumption within specific professions (Dyer 1988, 2000: 77-100). Documentation originating from

isolated communities such as Westminster Abbey (Harvey 1977) and the Durham Cathedral (Threlfall-Holmes 2005) similarly provides insight into practices of dietary and material consumption maintained by certain religious groups, however it is worth noting the non-prototypical nature of these cases relative to the general habits of consumption practised on a broader scale. Like modern investigations of consumerism, these studies predominantly focus on the person(s) doing said consuming, followed by a secondary focus on the processes through which this consumption is able to occur.

Recent work on medieval consumption has particularly benefitted from interdisciplinary approaches. Woolgar, Serjeantson, and Waldron (2005), for instance, combine textual and archaeological analysis in looking at dietary nutrition, a topic shared by Muldner and Richards (2005: 39-48) who explore the ability to reconstruct historical nutrient content through isotopic bone analysis. These new approaches, be they incorporated from the sciences or other humanities subjects, are powerful in their ability to independently confirm existing theories of consumption and typically provide further insight into an area of knowledge that would be difficult to otherwise obtain.

The present aim of exploring consumption in *PP* specifically focuses on dietary consumption. The *Middle English Dictionary* (hereafter, *MED*) identifies seven senses of *consumen* ‘consume’, each featuring demonstrable changes in meaning and use by the beginning of the fifteenth century. The majority of these senses, or uses, identify agents through which people were thought to be consumed by, including fire and heat (1), worms or vermin (3), as well as disease, fever, and old age (4, 6); there is also precedent for being consumed by one’s adversaries, and the swords wielded by them (2). These agents of consumption contribute to stylised depictions of death as caused by both illness and violence that recurringly feature in literary texts such as Chaucer’s

Canterbury Tales and Lydgate's *Troy Book*. The term also refers to consumption as a using-up process or action manifested through ingesting foodstuffs (7a) and expending non-edible goods or property (7b), which additionally extends to include expenditure of money or general wealth. In this sense *consumen* also has the capacity to operate figuratively through the representation of time as something to be consumed (7c).

Examination of the language used to represent dietary consumption in *PP* is intended to complement a corresponding focus observable in the text's critical scholarship. These studies tend to document textual examples describing processes of eating and drinking in the poem, then link these actions to overarching thematic strands. Mann (1979), for example, analyses these two bodily processes as textual embodiments of higher thematic concepts oriented around gluttony, social waste and, at times, the Eucharist. She observes the pervasiveness of these images across the text (1979: 26), and furthermore acknowledges the difficulty at times of separating the literal from the metaphorical in Langland's references to these acts of consumption. Schmidt (1980) also hones in on this use of allegorical imagery surrounding the representation of food and drink as evidence of spiritual disorder and corruption in specific sections of the poem, discussing the meaning of consumption in the text through the presence of allegorical characters such as Measure, Glutton(y), and Hunger. The significance attributed to the latter character is particularly contentious in Langland scholarship; Benson (2004: 138) and Gasse (2016: 67-8) summarise the debate as being an interpretative one pitting the literal significance of physical hunger as social critique against the more figurative understanding of hunger as a process of spiritual exegesis, thus building on Mann's (1979) initial observations on the dual interpretations rendered permissible through Langland's allegorical treatment of these consumption-based processes. These divisions in interpretative standpoint also treat

descriptions of dietary consumption as a physical manifestation of larger ideological challenges. Some critics characterise the text's approach to consumption as the struggle instigated by either internal or external forces (Savage 1993), while others frame it as a representation of the interface between warring Christian and feudal approaches (Kim 2004).

Taking a corpus approach to this theme offers a method of contributing an additional linguistic perspective to an already rich body of literary scholarship. Both the consumptive habits of the 14th and 15th century populace and the underlying agrarian and commercial processes responsible for meeting this demand are rigorously parsed and interpreted in both literary and non-literary source documents (i.e., charters and guild agreements) by existing historiographic scholarship, but little to date explores the formal techniques by which this information is linguistically conveyed and to what effect(s). A corpus approach enables these linguistic patterns to be better visualised by revealing how the the sense of dietary consumption described in (7a) of the *MED* (i.e., eating and drinking) are conveyed lexically in the poem, and what patterns correlate with their use. This focus on the lexical representation of dietary consumption is particularly useful in examining the structures of evaluation underyling the poem's evaluative negativity towards consumption through the semantic prosodies revealed in collocational analysis (see 2.3.4). A corpus linguistic approach also useful in determining whether the patterns depicting consumption are unique to *PP*, or are instead reflective of larger discoursal trends in language use within late medieval poetic literature.

2.1.6.2 Legality and the law

Unlike consumption, legality exists as both a linguistically-perpetuated ideological theme as well as a professional variety of language. The presence of language used primarily in legal discourse within other types of documents, including poetic texts such as *PP*, makes it more appropriate for the current investigation to refer to this specialised language as *legal language* (e.g. Tiersma 1999, Mooney 2014) rather than *the language of the law* (Bhatia 1987, Maley 1994) to address its roles outside of a strictly legislative or judicial functionality.[9] While legal language was originally approached by scholars such as Mellinkoff (1963: vi) as being 'a profession of words', the analytical scope of legal language has been broadened in more recent scholarship to consider the role of syntax and grammar in legal rhetoric (e.g. Gibbons 1994, Schane 2006, Mooney 2014). This has import for analysis of legal language in literature, allowing legal vocabulary and grammatical constructs represented in literary contexts to appear as recognisable, or 'fixed' features derived from use or association with the legal profession.

Legal language within a 14 th century context is fundamentally shaped by England's late medieval legal system, and the historical role of English within that system. Within this system legality was monitored and enforced with oversight from several overlapping benches adjudicating different aspects of daily life, with common law representing the primary legal system (Baker 2002). Brought about by the reforms set out by Henry II (1154-89), common law unified the country through a law-by-precedent model wherein the weight of past decisions was brought to bear on similar disputes to increase judicial standardisation. This system consolidated the power of

[9] In other words, legal language contained in literature is unlikely to play an integral role in upholding the law from a real-world perspective.

royal justice (the king's justice) into a central jurisdiction formed of two courts, the Court of Common Pleas and the Court of the King's Bench; the former addressed legal disputes not involving the king, while the latter focused on cases of criminality and those directly affecting or concerning the king (Holdsworth 1982: 29-38, Baker 2002: 14-22). Local structures of this centralised system were formed by the assizes in the mid 13th century, which outsourced investigations and trials from Whitehall to the local districts in which these disputes where initially raised through the help of additional local justices of the peace regularised under Edward III (Baker 2002: 20-6); these were supplemented by local civic officers like bailiffs and mayors which were chosen by the landed populace (Carrell 2007: 14-6). In opposition to common law, the other legal system operating in medieval England was canon law, which stood outside the remit of the secular courts and dealt with ecclesiastical matters stemming from the religious ordinances and regulations set out by the Catholic Church (Baker 2002: 126-30). The legal officers operating within this remit were members of the clergy including bishops, priests, and cardinals, some of whom also possessed a more formal legal education.

Within this legal landscape, the role of English was heavily influenced by vernacularisation in the post-Norman Conquest courts. English featured as one of three legal languages used in the courts, with English and Latin appearing in early written legal records and French taking a more heavily oral role alongside English in argumentation as well as court testimonies and verdicts (Mellinkoff 1963: 66-9). Of the three Latin was the most influential as the primary language of legal statutes, literature on the law, and other important documents until well into the 13th century, dominating the language of both Church and secular courts (Alford 1977) in a manner highlighted by the amount of Latinate legal vocabulary borrowed into the language

post-Conquest (Baugh & Cable 2002: 158). As the 14th century progressed English gradually replaced Latin (and later, French) as the primary language of the law, a movement started with the Statute of Pleading in 1342 asking for oral courtroom discourse to be changed to English for reasons of clarity and for the better protection of its citizens.[10] Despite initial reservations, particularly regarding the precedents and tradition established in linguistic legal formulae already present in French and Latin, English was increasingly associated with arguments at the bar, and and even in the composition of the statutes themselves by the end of the fifteenth century (Mellinkoff 1963: 111-8).[11]

Examining the presence of this legal language (English or otherwise) in poetic texts has led to some striking comparisons between historical legal and literary discourses. Scholars such as Steiner (2003) and Green (1999: 407) have argued that medieval legal and literary texts are more closely aligned than their modern counterparts to the extent where they are best studied 'as parallel forms of discourse, each with its own conventions and traditions'. Both medieval legal and poetic documents, for example, feature high amounts of lexical repetition – as well as grammatical coordination and parallelism – which reveal their shared history as discourses stemming from traditions of oral transmission. These lexico-grammatical similarities also correspond with similarities in overarching aim, with both legal and poetic texts demonstrating mutual interest in contributing to the shaping of cultural ideals, ethics, morals, and values within the society of their creation, albeit through different means (Pollock & Maitland 1895, cited in Alford 1977). This stance

[10] This statute was, ironically enough, composed in French (see Mellinkoff 1963: 112).
[11] It is obligatory here to note that the described shifts between Latin, French, and English in late medieval legal language occurred gradually. There was a great deal of code-switching between languages during transitional periods; additionally, most surviving legal texts demonstrate bilingual characteristics (i.e., formulaic Latin phrases surviving in early 14th century French documents).

corroborates Alford (1977: 451-2), who argues that 'the association between literature and the law has never been more impressive…than in Medieval England'. Steiner and Barrington (2002: 11) further attribute this connection between the law and texts like *PP* to a broader cultural movement of linguistic and ideological reassertion in the 14th century, claiming that

> the slow demise of Anglo-Norman as a discourse and a practice, and its eventual marginalization – if not diminution – in the written record, offered late medieval English poets a space to recuperate the relationship between the legal and the imaginative…[giving] Middle English writers the right to respond to legal practices and reshape legal discourse.

What separates legal language in conventional legal discourse from its appearance in literature is its textual purpose. While both utilise the same knowledge schemata concerning the late medieval legal system and its components, its presence in medieval literature provides a fictional representation of real-world law which does not have to be factually accurate or realistic in a real-world sense. Legal language in poetry does not fulfil a real-world judicial or legislative function, freeing it from the careful regulatory framing needed to ensure universality covering general and individual circumstances and the need to maintain flexibility in the manifold and complex responsibilities required from the words constituting the law (Crystal 2003: 374). Language in legal discourse aims for universal definition and strict adherence to precision, often achieved through recourse to established rhetoric and conventionally-tested linguistic frameworks and phrases (Tiersma 1999: 95-7). By comparison, poetic language seeks to depart from established tradition and stretch to its pragmatic and metaphoric limits by bending and otherwise altering meanings for creative effect (Alford 1988a: xiii); this approach often creates ambiguity where connections are either obliquely referenced or completely unvoiced (Green 1999: xvi). This creative

potential in exploiting the ambiguity actively minimalised in legal discourse makes the poetic use of legal language in *PP* a topic of interest from a stylistic perspective.

The most comprehensive study of the legal language in *PP* to date lies in the glossary of legal diction compiled by John Alford (1988a). Based on the premise that *PP* contains unusually high amounts of legal terminology – a stance of universal consensus in Langlandian scholarship – Alford (1988a) lists the over 700 legal terms found in the text (most of them used more than once) and reveals over 100 legal concepts for which the poem provides the earliest recorded testimony in English (Alford 1988a: vii-ix). He also provides legal definitions for terms drawn from English, Latin, and French as they were used in a late 14th centry context. Pearsall (1988: 363) highlights the value of Alford's (1988a) work as including, in addition to highly technical legal terminology, the 'technical legal uses of what look at first sight to be familiar words (e.g., *countrees, paper*)'. These entries, containing terms whose earlier legal connotations have been eroded or completely erased over time, are of particular use to modern scholars. While not providing a listing of all occurrences of each legal term in the text – perhaps on account of a *Piers Plowman* concordance not becoming available until after its publication[12] – the glossary is a first of its kind, offering not only a solid point of departure with which to begin looking at the poem's use of legal terminology but also a template for investigating the legal language used by other late medieval literary authors.[13]

In addition to Alford (1988a), interest in the poem's legal language – and their thematic import to the text – represents a niche area of *PP* scholarship. Literary critics have examined its textual presence through addressing instances of legal metaphor

[12] See Wittig (2001).
[13] Barrington (2010) uses Alford (1988a) to look at legal language in Gower, while Barr (1994:133-66) uses it as a baseline in exploring similarities in legal diction within the anonymous poems *Pierce the Ploughman's Crede, Richard the Redeless, Mum and the Sothsegger,* and *The Crowned King*.

(Birnes 1975), types of legal argumentation (Stokes 1984), thematic (Baldwin 1981) and metaphoric representations (Alford 1988b), and even legal satire (Simpson 1990). This focus on legal language has also fuelled speculation regarding the author's professional background as a minor clerk (Kirk 1933: 237) and even been used in authorship attribution studies as a method of distinguishing between scribal voices (222-5); however, most scholarship tends to take a historiographic approach by comparing how the legislative minutiae evidenced within the text is reflective of its late 14th/early 15th century context. More recent studies have returned to the proximity evidenced between the legal and the literary initially identified by Alford (1977), with focus oriented around how the processes of informing and textually transforming caused in literary mediums like *PP* creatively alter legal significance to achieve authority in other domains (Thomas 2012) and even engenders its own poetics (Thomas 2018). This research has collectively drawn attention to the import of legal language in *PP* not only through the sheer textual density of these terms but also through their influence on the structure of the text, 'whose mode of thought and argument is substantially indebted to legal procedure and vocabulary' (Barr 1994: 133).

The present study offers a fundamentally different approach to legal language in *PP*. The ability of corpus tools to show a word in its immediate context enables observation of all the demonstrated uses of that node word through its surrounding textual patterns (Scott & Tribble 2006: 33). This contextualisation is particularly important for medieval legal terminology, of which 'many, especially those with English roots, are visible only in context alone, particularly the way in which they are joined with other words' (Barrington 2010: 115, citing Alford 1988a). This relationship between pattern and function can facilitate distinction between legal and non-legal

uses in polysemous and homographic terminology, as well as distinguish between various legal applications of these terms as stylistic choices within legal diction. It can also prove useful in identifying extended units of meaning (Mahlberg 2013: 15), especially those terms which, when viewed as contextually isolated elements, do not retain their legal significance.

More importantly, corpus methods assist in identifying the significance of the textual presence of this legal terminology relative to its use in other medieval literary texts. Addressing the comparative use of these legal terms within (and across) different poems offers a different type of empirical information than gathered by counting raw frequency of occurrence within the text, and fills a gap in current legal *PP* scholarship; many studies look at the role(s) of legal terminology within the poem or different versions of the poem (i.e., Kirk 1933), but very few offer any basis of comparison to the surrounding literature. Barr's (1994) chapter on the use of legal language across the *Piers Plowman* tradition is a notable exception to this, but offers only qualitative grounds for comparison rather than a combined qualitative and quantitative study. To this end, conducting a keyword analysis makes it possible to identify the legal terms that appear in *PP* with greater-than expected frequency (i.e. statistical atypicality) compared to other late medieval poetic texts on a quantitative basis (see 2.3.3). A corpus linguistic approach also offers the opportunity to examine the strength of affiliation between identified legal terms to other words appearing in their textual proximity via collocational analysis (see 2.3.4) which can reveal meaning of stylistic significance, particularly through the creation and relative impact of metaphoric ambiguity.

The following section (2.2) takes the poem and situates the stylistic examination being conducted in the present study within existing scholarship and

methodological traditions within literary linguistics, corpus linguistics, and corpus stylistics.

2.2 Literary linguistics

At the most fundamental level, literary linguistics (or *literary stylistics*) is concerned with the study of literature through a text's linguistic features. A method of textual interpretation, it stands apart from conventional literary criticism through the assignation of language as the primary feature of interest (Simpson 2014: 3). This linguistic emphasis enables observation of the text's intra-linguistic features and the manner in which these features interact to answer not only what a text means, but how that meaning is generated (Leech & Short 2007: 2).

The current investigation operates within the definition of literary stylistics set forth by Semino and Culpeper (2011), who follow Leech (1985) in distinguishing *literary stylistics* as the study of literary texts, and *general stylistics* for other discourse types. Within this narrowed scope of literature-based study, Wales (2001: 373) attributes a greater literary focus to the term compared to the more language-oriented focus given to *linguistic stylistics*. However, my aim to study language use as well as literary effect situates my research in both spheres equally; given this, I instead follow Carter's (2014: 84) stance observing the declining need for such a distinction. While Stockwell and Whitely (2014: 4) acknowledge the field's referential variance according to geographic region (including but not limited to: *literary stylistics, literary pragmatics, poetics, literary discourse analysis*, and *stylistics*), my decision to use *literary linguistics* in conjuction with *literary stylistics* is meant to offset potential confusion by clearly distinguishing the current approach from medieval literary

traditions associated with terms such as *style* and *poetics* (see 2.2.1). Contemporary literary linguistic study retains historical roots in *rhetoric* which also features in critical literary analytical frameworks (Wales 2014a: 668-9), but has since diverged from its classical aims and in the process embraced new analytical theories (Hamilton 2014, Burke 2014).

The purpose of this section (2.2) is not to provide a detailed discussion of literary linguistics (hereafter, LL), as this has already been comprehensively addressed by existing scholarship. Recently published handbooks (Lambrou and Stockwell 2010, Stockwell and Whiteley 2014, Burke 2014, Sotirova 2015) offer a broad introduction to the various language features, techniques, core issues, contextual applications, contemporary topics, and emerging trends associated with the field, with resources such as Wales (2001, 2014a) and Norgaard et al. (2010) defining key terms from a LL-based perspective. This research builds on core scholarship in LL including Short (1996), Leech and Short (1981, 2007), Verdonk (2002, 2013), Leech (2008), Simpson (2004, 2014) and Jeffries and McIntyre (2010) which collectively feature (and in doing so, perpetuate) the 'stylistics toolkit' metaphor to characterise the field's focus on certain linguistic features and patterns as cues for examining stylistic effect (see Wales 2014b: 32-3).

This section instead addresses the basic tenets of style in a modern LL context so as to distinguish it from the aims of stylistic examinations conducted in medieval criticism and within other linguistic spheres (2.2.1). It also addresses some key aspects of approaching style within a historical context (2.2.2), as well as the features of late medieval poetic language making a corpus approach to style useful to the present investigation (2.2.3).

2.2.1 What is literary style?

Central to examining *PP* stylistically is first defining what constitutes literary style. For this investigation my use of the term *style* draws on the dual broad and narrow conceptual parameters outlined by Leech and Short (2007). A more narrow interpretation of style focuses on its textual features as the 'linguistic choices which concern *alternative ways of rendering the same subject matter*' (Leech & Short 2007: 31, original emphasis). This view bases the creation of style on authorial choice based on the distinction between 'what is a*ctually written* with what *might have been written*' from a total repertoire of given language (Short 1996: 68, Short's emphasis). Leech and Short (2007: 31-2) additionally offer a broader definition acknowledging style as being both situated and context-dependent, with said choices operating within 'a domain of language use' containing conventions characterising a particular text or literary tradition. In addition to providing extra-linguistic contextualisation, this broader conceptualisation of style ascribes a duality within language in literature[14] as being functional as well as aesthetic.

This distinction between aesthetics and function is not clearly delineated, particularly when exploring the language contained in poetry. The language appearing in poetry is unusually representative of linguistic multi-functionality in its use; unlike prose and dramatic language, it has to conform not only to the grammatical rules

[14] My phrasing 'language in literature' is meant to disambiguate my reference to the language present in literary texts from the concept of 'literary language' as being a variety of language found exclusively in literary contexts as estranged from everyday language use (for more, see Simpson 2014: 106-100). For the purposes of my research I use 'literary language' as synonymous to 'language in literature', and adopt Carter's 'cline of literariness' (2004: 66) to address the more stylised use of language observed in literature compared to other discourse types.

governing the content of the poem but also to the structural limitations in form imposed by its metrical requirements (Short 1996: 134). These factors produce a large number of 'compact' (Crystal 1987: 71) textual forms performing both aesthetic and functional roles simultaneously within the poem.[15] Such linguistic forms – including patterns utilised in *PP* such as alliteration and personification allegory – are dependent on poetic creativity from the writer to achieve a level of textual sophistication without sacrificing the poem's overall coherency.

To better discuss the aesthetic and functional roles played by the language appearing in *PP* I draw on Biber and Conrad's (2009) distinction between *register* and *style*. This differentiation frames linguistic features or patterns characteristic of (and having an established function in) a given set of text-types[16] as being indicative of register, while features less directly functional in their aesthetic presence are more stylistic:

> Register features are pervasive linguistic features that are functional; that is they are frequent because they conform to the situational context and communicative purposes of the texts in the register. Style features are similarly pervasive linguistic features but they are not directly functional. Rather, they reflect attitudes about language, and aesthetic or artistic preferences. Thus, texts from the same register, sharing the same situational context and the same communicative purposes, can differ in their linguistic *styles* (Biber & Conrad 2009: 151, original emphasis).

In this sense, the alliteration observed in *PP* broadly operates as a register feature of Alliterative Revival poetry but the atypical intra-linear alliterative pairing of function and content words discussed by Robertson and Shepherd (2006) would be more representative of a style feature.

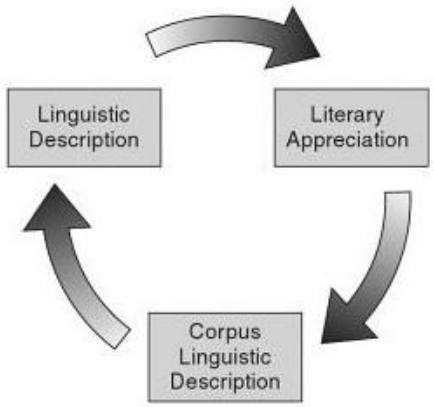
---

[15] This is particularly visible in instances of poetic metaphor (see Briggs & Monaco 1974; Brouwer 2010).
[16] I adopt Culpeper and Kyto's (2010, citing Taavitsainen, 2001: 140) use of the term *text-type* to differentiate discussion of a specific, textually-based type of fictional discourse from the cultural elements of production, distribution, and consumption associated with *genre*.

A crucial component of stylistic expression lies in literary appreciation, or the reader-based perception of style (Leech & Short 2007: 11). While literary style is grounded in the language contained in a text or set of texts relative to its surrounding linguistic context, its expression only takes on significance through the identification and subsequent recognition of said choices as being 'stylistic' by its readers. This aspect of style is more closely examined by scholars in in cognitive poetics (for example, Stockwell 2005: 747-8) who signpost the importance of textual reception by observing that style does not exist as an inherent element within a text, but manifests itself in the interaction between text and reader. Readings of historical texts like *PP* would ostensibly vary in the types of literary appreciation derived from exposure to a modern or contemporary medieval readership; the appreciation generated from modern readings would also fluctuate according to individual experiential familiarity with a) the text and its content, b) Middle English (if untranslated), and c) the historical and literary traditions and conventions surrounding late medieval narrative English poetry.

This sense of literary style accessible to the reader as effect is intrinsically tied to the perceived literary worth of the text in a cyclical process linking textual form – as manifested in its linguistic elements – to literary appreciation. This process is illustrated by Leech and Short's philological cycle (2007: 11-2, citing Spitzer 1948) which Mahlberg (2013: 12) updates with the inclusion of corpus methods and theory (for more discussion, see 2.3.1). This process, as represented in Figure 2.2, typifies the investigative approach on which my present research is modelled.

Figure 2.2: The corpus stylistic circle, as outlined in Mahlberg (2013: 11-2)



This LL focus on style as the relationship between linguistic form and resulting literary effect diverges from the approaches characterising other modern linguistic fields, particularly authorship attribution (hereafter, AA). From an AA perspective, style corresponds to the perhaps more intuitive ideation of the 'linguistic fingerprint' belonging to an individual author (i.e., Hoover 2007: 176), with research dedicated to compiling lists of distinctive linguistic features characterising the author's writing by examining 'linguistic traits which may not necessarily be artistically relevant… [and] beyond conscious artistic control' (Leech & Short 2007: 12). AA constitutes an area of interest for researchers studying historical texts (including but not limited to: Horton 1987, Burrows 2002, Hoover 2004, Burrows 2005, and Craig & Kinney 2009), and is visible in the approach to *PP* evidenced in early twentieth century Langlandian scholarship within studies championing theories of single versus multiple authorship. I use Fialho and Zyngier's (2014: 330) term *stylometry* to distinguish LL-based stylistics from these types of authorially-oriented linguistic examinations of style.[17]

[17] A list of key AA-based studies (some including stylomtric analysis, though not signposted as such) for *PP* can be found in Pearsall (1990: 45-64); for a general summation of AA studies of *PP*, see Kane (1965).

In addition to modern linguistic distinctions in usage *style* also conveys different meanings within a medieval critical context. These uses can broadly be condensed into two approaches, with the first representing style as 'authorial personality'. Hughes (1995: 34) acknowledges the popularity of this approach as 'a potent medieval method of defining style [through] the creation of a fictional psychological profile of the writer' which, for texts like *PP*, are 'concerned not so much with describing Langland's style nor explaining its technical components as with tracing style to its efficient cause in the poet' (Davlin 1995: 31). This approach is exemplified by scholarship like Barnes (1995) who, Hughes (1995: 35) argues, 'offer[s] a psychological profile of the sort of person Langland must have been to have written in the style he did – "being the kind of person he was…"'. Unlike the aims of LL (which tend to eschew knowledge of the author during language analysis), these investigations actively incorporate external knowledge of the author and their lives as identifiable historical figures to produce these profiles; in the case of anonymous texts or ones with tenuous attribution histories like *PP*, knowledge gaps about these authors are backfilled with information drawn from the texts themselves. Schmidt (1987), for example, uses the professional domains evidenced in *PP*'s lexis as the basis for further extrapolation about the author as someone possessing knowledge of contemporary legal affairs as well as poetic conventions as a 'clerkly maker'.

The second medieval-based approach uses *style* to refer to a list of features attributable to (or characteristic of) a literary author, text, or greater literary period. Synonymous with the more classical referents *poetics* (Wales 2014a: 323), this approach is typically characterised by a deconstructionalist view focusing on the text and its features over historical content and authorial study to highlight some of the language features behind the perceived distinctiveness of a given text. In this sense

'style as poetics' shares analytical interests with AA (i.e., interest in the link between language patterns and individual texts/authors), but stands apart in that it focuses on the language patterns present in texts whose authors have already been identified. Examples of medieval literary scholarship fitting within this category include studies on the language of Chaucer (Horobin 2012) versus Gower (Smith & Samuel 1981) and Langland (Rudd 1994).[18] Poetics also stands apart from LL in the view that the language evidenced in poetry is fundamentally more 'literary' than other types of language use (2014a: 323).

This 'style as poetics' approach is best exemplified by Lawlor's (1962) examination of *PP*'s language wherein the poem is broken down into 'argument' (i.e., content) and poetic techniques, with the latter covering features including but not limited to metre, rhythm, allegory, and wordplay. While a good part of his discussion is spent listing features of the text's poetic language, some of his observations are things that resonate with contemporary LL analysis, particularly in the significance of recurring text patterns.[19] His work evidences recognition of the significance of non-divisible multi-word lexical units in the poem (as distinct from the pre-set alliterative pairings present in other poems) (1962: 195, 201), and also signposts the lexico-phonetic parallelism produced in alliterative pairings (1962: 190) as well as the conceptual foregrounding resulting from the combined effects of metrical stresses falling on content words (1962: 191). While perhaps not using the same terminology recognised in LL, some of the effects being discussed are comparable, despite

[18] Some studies do look at the language of 14th century poets more generally; for examples, see Smith (1988) and Horobin (2003). These studies effectively extend the scope of poetics from authors to literary movements, which in this investigation is distinguished by the style vs. register distinction.
[19] While examining *Sir Gawain and the Green Knight* rather than *Piers Plowman*, Borroff (1962) represents a similarly seminal work bridging poetics with modern literary linguistic analysis within a medieval literary context.

differences in view regarding the literariness embedded in poetic language behaviour and use (see also Donner 1987).

These four approaches to medieval literary style – LL, AA, 'style as authorial personality', and poetics – all draw on linguistic features for greater literary understanding (Wales 2014a: 400) but with fundamentally different aims. These diverging aims are illustrated below by indicating how each field would approach the alliteration contained in *PP*.

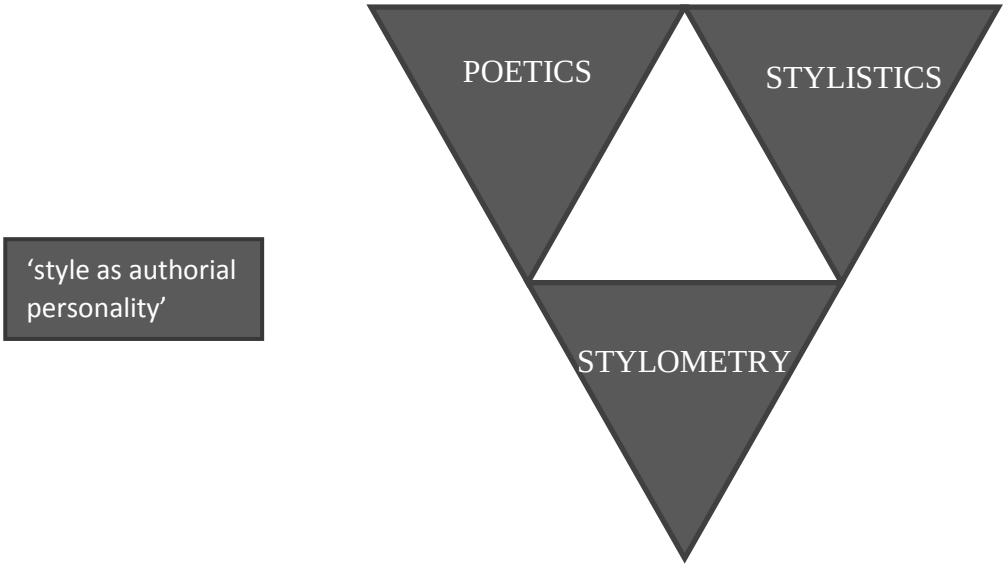
> **AA** – are the alliterative patterns in *PP* observable in any other contemporary texts (i.e., could Langland have written some of the other presently anonymous poems in the surviving late medieval literary canon).
>
> **Poetics** – what types of rhythm and stress patterns are characteristic of the text's alliterative metre.
>
> **Style as authorial personality** – what does the use of alliteration in *PP* reveal about the author.
>
> **LL** – what literary effects are generated from the use of surface-level alliterative structural patterns (i.e., foregrounding through repetition and phono-syntactic parallelism).

From amongst these approaches *PP* scholarship predominantly falls under the AA and 'style as authorial personality' categories, with a scattering of works conducting analysis reminiscent of poetics (i.e., Lawlor 1962). The closest example of a directly LL approach lies in the work conducted by Muscantine (1999) who, while predominantly focusing on Chaucer (i.e., Muscatine 1972), applied the modern stylistic theories stemming from Auerbach and Spitzer to medieval poetic literature more generally.[20] Further distinction between these analytical domains is illustrated in the diagram contained in Figure 2.3, which demarcates stylistics (LL), stylometry (eg. AA), and poetics as the primary approaches to style, with 'style as authorial personality'

[20] Davlin (2018) offers a good overview of Muscatine's influence in incorporating a more literary stylistic approach into the study of 14th century literature.

operating as a hybrid field integrating aspects of poetics and stylometry simultaneously.

Figure 2.3: The different approaches to *style* characteristic of modern applied linguistic and medieval literary critical research (updated from version presented in Cordell & Wilkins 2017).



In distinguishing between these perspectives I situate my research within a stylistic approach, wherein emphasis is placed on the literary effects generated from use of recurring textual patterns as a method of critically engaging with the poem. In this capacity my view on producing a list of distinctly 'Langlandian' features corresponds with Mahlberg's study (2013: 24) in that the value of such a list 'would be doubtful in a [stylistic] framework that includes literary appreciation as one of its key aspects'. While a list of features favoured by Langland may become evident during the the project, they do not represent my primary investigative aim. My study instead focuses on the relationship between the poem's linguistic features and the effects they produce during the interpretative process performed whilst reading, with the author's

choice of language use representing one of several factors contributing to the creation of perceived style.

2.2.2 (New) historical stylistics

Historical literary linguistics, or *historical stylistics*, refers to the use of a LL approach as applied to historical literature (Norgaard et al. 2010: 26-7). For the purposes of the present investigation I define 'historical literature' as works composed through the 19th century inclusive. Delineating literary historicity here as pre-20th century is intended to reflect existing trends in LL scholarship, which heavily favours investigations of contemporary literature and language (eg. Short 2008: 3). The past decade has revealed increasing numbers of LL studies examining progressively more historical texts and authors including novels by Austen (Bray 2010, Massimiliano 2010, Fischer-Starke 2010) and Dickens (eg. Mahlberg 2010, 2013), Keats' poetry (Giovanelli 2013), and even Shakespearian drama (eg. Busse 2010a, Bousfield 2010, Leech 2010, Ravassat & Culpeper 2011).

The present investigation follows the theoretical framework for studying historical style set out by Busse (2010b, 2014). Her label '(New) Historical Stylistics' (hereafter, N-HS) signals 'a new diachronic turn in stylistics', which identifies the emergent trend focusing on the application of stylistic tools, methods and approaches to historical literature (2010b: 33-4). Its utilisation of recent developments in modern literary linguistic analysis offer new methods of engaging with historical texts and literature which, in addition to providing novel ways of comparing and linking language data, also allow new types of questions to be asked (Busse 2014: 102). My use of the term N-HS here broadly addresses studies characterised by this approach to the

language appearing in historical literary texts, whether or not actively signposted as such within the works themselves.

The primary tenets of N-HS are based on awareness of the contextual surroundings and normative processes in which these historical texts are situationally embedded. In this sense a N-HS approach 'presupposes a comprehensive knowledge of the period, the context, and the language in which the text was produced', with context determining linguistic function and import (Busse 2010b: 35). This approach seeks to replicate the analytical detail and wider contextual linguistic and extra-linguistic factors influencing historical language use characteristically scrutinised in traditional historical linguistic analysis, but diverges from classical linguistic areas such as phonology or morpho-syntactic structures in its focus on features and patterns contributing to literary effect (Busse 2010b: 33). These underlying tenets mirror the methods and values adopted in other contemporary linguistic fields examining historical language use including historical sociolinguistics (Anipa 2012, Auer et al. 2015), historical pragmatics (Taavitsainen and Fitzmaurice 2007, Fitzmaurice 2010), and even the representation of historical dialogue (Culpeper and Kyto 2010). Busse (2010b: 40-1) provides a Short-inspired 'checklist' for N-HS outlining its key points and general analytical aims.

The most prolific clustering of historical LL investigation appears in the research surrounding the language of Shakespeare and his Early Modern English contemporaries. Examples of these include pragma-stylistic investigations (eg. Bousfield 2010, Archer & Bousfield 2010), corpus-based studies (Culpeper 2009, Archer & Bousfield 2010, Anderson & Crossley 2011, Demmen 2012, Murphy 2015), and research adopting methods and frameworks drawn from cognitive poetics (Cook 2011, Oncins-Martinez 2011, Sprang 2017) and characterisation studies (Culpeper 2001,

2002, 2009). While not a substantive body of scholarship compared to the amount of research conducted on Modern English texts, it currently represents the largest clustering of historical literary stylistic analysis to date. The present exploration of *PP* as a late medieval text is meant to assist in shifting the current state of historical LL diachronically from studying EME to the late ME operating within an earlier period of English literature.

This is not to claim that historical stylistic approaches are not currently being applied to texts more historical than Shakespeare, merely that such investigations are comparatively rare. While Pons-Sanz (2015) provides a poetics-based study of OE style by outlining the primary features evidenced in OE prose and verse and how the Norman Conquest altered the use of these patterns (eg. replacing alliteration with syllabic rhyme as structural devices, and an emerging preference for loanwords over poetic compounds), Harbus (2006) represents one of the first clearly identifiable modern literary linguistic examinations of pre-Shakespearian literature. Her work (2006, 2016) actively incorporates and engages with modern stylistic theories such as conceptual metaphor (Lakoff and Turner 1989, Lakoff and Johnson 1980a), blending (Fauconnier and Turner 2002), and text world theory (Gavins 2007) into analysis of a selection of Old English poems and riddles. Using cognitive stylistic frameworks to reveal new insights about language use and meaning in OE literature has been more recently addressed in Wilkins (2018), which examines explicit cues (Culpeper 2001), deixis (eg. Green 2008), and discourse presentation in modern interpretations of Aelfric's *Esther*.

The value of this research from a N-HS perspective (and, by extension, LL more generally) lies in its reflective discussion of the similarities and differences embedded in these theoretical frameworks when applied to very historical material. For example,

Harbus (2006: 89-94) demonstrates how text-world theory can be used to pinpoint linguistic triggers that signpost the nested mental spaces characteristic of OE poetry that 'make strenuous cognitive demands on its audiences' by making users hold multiple levels of levels of intentionality in mind simultaneously during stages of textual processing (2006: 178-9). This not only reflects the durability of theories of mind-reading (eg. Zunshine 2003, 2006) in distant historical literary contexts, but also reveals new methods by which stylistic features like negation and megametaphors are textually expressed, especially via techniques not conventionally found in modern literary contexts.

In addition to Old English, LL-based approaches are also to some degree visible in the scholarship surrounding Middle English literature. While appearing before recent advances in literary linguistics (and thus not described as such in approach), critical literary studies studies such as Borroff (1962), Lawlor (1962), Muscatine (1972, 1999), and Wittig (1978) evidence similar interest 'in the development of linguistic models of analysis and their application to both stylistic and narrative study' (Wittig 1978: 3) through their poetics-based foci. Collectively addressing examples drawn from the late ME literary canon, these studies intermittently echo modern LL in the examination of 'characteristic stylistic and structural redundance', the production of parallelism and prominence as effect in the minds of ME poetic audiences based on said patterns, the need to observe works within their greater generic contexts to adequately signpost stylistic atypicality, and the presence of sets of structuralist-based hierarchical systems distinguishing between 'small-scale stylistic forms' versus 'patterned groups' and broader 'formulaic plot structures' (Wittig 1978: 3-9, Lawlor 1962: 190-201, 229-33). These critical literary studies are augmented by parallel approaches visible in historical pragmatic explorations of late ME literature (eg.

Pakkala-Weckstrom 2010). Sylvester (2015) provides the first overview of ME literary discourse from a modern LL perspective, with upcoming research starting to break into the late ME literary stylistic sphere through Collins' (forthcoming) examination of textual cohesion, intertextuality, and framing devices in Malory's *Morte Arthure*.

The present investigation studying *PP* is thus situated within a new wave of LL exploration which, assisted by existing language-oriented critical scholarship, focuses on the application of modern literary stylistic approaches and theory to pre-Shakespearian texts and language. Stylistic approaches to *PP* in particular necessarily operate within the context of the poem's presence in multiple textual versions, with its clear chronological progression of revision illustrated in the ABC cycle indicating a conscious interest in style during the writing process as demonstrated by the author 'working his way towards a particularly appropriate and effective set of linguistic choices' (Short 1996: 69-70). While I generally aim to steer clear of discussions concerning authorial intentionality, the composition and subsequent alteration of the text over the course of a twenty-five year period from the early 1360s to 1387 in 'a prolonged and continuous act of composition' (Benson 2004: 43, citing Chambers & Grattan 1909) makes it no great stretch to attribute a broader stylistic design to the poem – in its revision, if not its original reason(s) for being written.

2.2.3 Style in Middle English poetry

Studying *PP* compared to the language evidenced in other late medieval poetic literature situates the present study within a specific set of contextual parameters. The use of poetry focuses my analysis on a specific text-type, a decision made in efforts to sidestep issues of what distinguishes 'literary' from 'non-literary' texts written in

Middle English by basing selection on poetic form (see Burrow 2008: 13-24). Within this type of literary discourse, each of the 28 texts selected for use are thought to have been composed within a designated time period (late 14th/early 15th century) broadly following Brown's (2009) 1350-1500 timeframe demarcating the late medieval English literary movement to which *PP* belongs. This temporal specification subsequently narrows analytical focus to a synchronically approximate language form (non-regionally specific late Middle English).

Narrowing the domain of language use to Middle English poetry composed within (or around) the late 14th century thus orients this investigation of style within a pre-established linguistic sphere. This approach is intended to produce a representative set of relatively homogenized linguistic data, or a specific 'variety' of language (Crystal 1970: 101) reflecting greater overall similarity than difference meant to characterise the conventions of language use evidenced within late 14th and early 15th century medieval poetry. Such conventions – encompassing the linguistic, cultural, social, and ideological factors (Fowler 1981: 174) influencing the production of late 14th and early 15th century language use – create an approximation of a language 'norm' for late medieval English poetry from which the similarity and difference to language use in *PP* can be comparatively evaluated. I say 'approximation' here intentionally to acknowledge the difficulty surrounding how to define a language norm – and if defined, the significance therein (see Wales 2014a: 292-294).[21] Existing historical LL analyses including Fischer-Starke (2010) and Mahlberg (2013) illustrate the benefits of comparing sets of historically contemporary, discourse-specific conventional language through the acknowledgement of style as something revealed

[21] The debate around this issue lies not in the underlying notion that style involves engagement with (and to some degree, departure from) expected language use within a given context (Simpson 2004: 3), but rather how this concept of normative language is defined, what it includes, and how to evaluate language use based on these parameters.

through knowledge of (or access to) the dominant conventions of literary form within the relevant time period (in both cases, 19th century novels).

However, there are some issues with ascribing linguistic conventionality or normativity to the language appearing in late Middle English poetry. For one thing, the 28 poems used in the current study do not constitute an exhaustive listing of the poetry thought to have been composed between 1300 and 1500. In this sense the language contained within these texts cannot be claimed as being fully representative of late medieval English poetry (for more discussion, see 3.1.1). Linguistic normativity also raises some problems based on variation in poetic form, given the presence of two separate poetic literary traditions operating within a late medieval context (i.e., the occurrence of the Alliterative Revival concurrent to a parallel tradition of metrical poetry).

The presence of these independent poetic traditions ultimately results in the production of two separate language norms operating within ME poetry, with each retaining their own standards of conventional language use tailored to each tradition.[22] Burrow (1971: 11-12) identifies these poetic language norms as the basis for underlying issues surrounding a normative conception of poetic style for the time period, claiming that 'the coexistence in the period of two such different traditions of verse-writing as the alliterative and the non-alliterative or 'metrical' told against the realizing for any common stylistic ideal, even supposing that such an ideal could have been formed by poets who were not, generally speaking, in communication with one another'. Burrow (2008) further reveals the problems with 14th century stylistic

[22] For example, (Sylvester 2015: 585) points out the greater use of ON borrowings in alliterative poetry, particularly within Northern or North-Midlands regional texts.

conventionality in the lack of an attributed name (such as 'Victorian' or 'Romantic' literature) for the given literary period. Such terms, he argues,

> are pregnant with a certain elementary but indispensable sense of the time period. In a literary context they suggest certain norms – of style, theme, convention, philosophy, etc. – to which even the most fastidious and sceptical critics refer…in the case of Chaucer and his contemporaries, the sense of such norms is lacking.'(2008: 1) [23]

The presence of multiple sets of conventional language in late medieval poetry thus situates discussion of normative language within (rather than around) a single text-type, and merits consideration of language use according to parameters beyond manual observation of poetic convention. How these two identified language norms are addressed in the current study is further examined in 3.1.1.

Another issue with a modern approach to language norms in ME poetry lies in the cognitive distance between text- and reader-based discourse worlds (Gavins 2007: 9-10). The temporal divide between *PP*'s date of composition and its modern 20th/21st century readership creates difficulties identifying the linguistic features and patterns belonging to the language norms characteristic of the time period, issues that would not be expected to be problematic for contemporary medieval readers: 'as modern historical linguists, we can only reconstruct the past, because we are not witnesses to the past usage of the English language' (Taavitsainen & Fitzmaurice 2007: 11). This lack of intuitive knowledge about what constitutes conventional language use in Middle English poetry makes identifying instances of linguistic irregularly or deviation problematic, especially given that these features in many cases signpost patterns of stylistic import (Short 1996: 10-13, Leech & Short 2007: 39-46). Medieval literary critics

[23] He offers the term 'Ricardian' for the literary period in recognition that the majority of texts from this time period were composed around the reign of King Richard II from 1377-1399; however, this term is meant to broadly address all literature composed during this period, not just poetry.

studying style have predominantly circumvented these intuition-based difficulties by acquiring experiential knowledge through dedicated study of *PP* and other late medieval poetry, operating as a type of cognitive reconstruction enabling better recognition of recurring linguistic patterns and conventions. While useful and resulting in a corpus of nuanced literary scholarship, this approach does take time and results in analysis that can be difficult to replicate based on the varying degrees of experiential knowledge developed by the individual researcher, as well as the levels possessed by persons evaluating said research.

My approach to normative language use draws on advances in N-HS and corpus linguistics to define the parameters of linguistic conventionality. In defining 'norm' I keep with the general stylistic practice of eschewing its prescriptive implications for a more descriptive-based significance (e.g. Stockwell 2005: 744), and follow the tradition set by scholars such as Busse (2010: 35-7) and Taavitsainen and Fitzmaurice (2007: 21) in broadly conceptualising normative language from a historical perspective as a range of language use that would reasonably be expected to be seen in a late medieval poem by its contemporary readership, as well as educated modern readers familiar with the literary conventions and linguistic expressions typical of the time period. In addition to this N-HS ideation of normative language, a corpus approach identifies linguistic conventionality through replicable, empirically-derived statistical parameters of language use developed in corpus linguistics, as set out in more detail below.

2.3 A corpus approach to style

Taking a computational approach situates the present study within a corpus linguistic framework. Broadly defined as the automated study of language through examples of

'real life' language use (McEnery & Wilson 2001: 1), corpus linguistics facilitates the examination of language through the compilation and analysis of corpora (Kennedy 1998: 1), or collections of machine-readable texts stored in electronic databases (Baker et al. 2006: 48). The linguistic content within these corpora (or corpus in its singular form) is meant to represent a sample of language use (Sinclair 1996) intentionally collected according to set linguistic criteria 'external to the texts themselves and dependent upon the intended use of the corpus' (McEnery et al. 2006: 4).

My aim to study *PP* within its immediate literary context situates this investigation at the smaller end of the corpus linguistic spectrum. Mahlberg and Smith (2010: 449) observe that corpora are prototypically large, with modern web-based corpora frequently surpassing million and even billion-word markers. *PP,* while a long narrative poem by English literary standards containing over 70,000 words across 7,338 lines of text,[24] constitutes a fairly small corpus by comparison; however, this is by no means unusual in the application of corpus techniques to literary texts (Mahlberg & Smith 2010, McEnery & Wilson 2001: 117). These relatively small specialised literary corpora tend to involve the application of corpus methods to specific text-types in order to answer a specific set of research questions (Baker 2006: 26). Such corpora often feature a single literary text, generally chosen for its perceived cultural or literary value (Stubbs 2005: 6, Fischer-Starke 2010: 5).

This corpus linguistic approach to Middle English literature stands apart from the digitally-assisted methodologies utilised in traditional medieval literary criticism. Specifically, it eschews already existing large-scale corpora in preference for specialised self-built corpora tailored to the research aims and scope of the present

[24] Length of the C-text, the longest version of the poem, as set out by Pearsall (1994: 10). It is worth pointing out, however, that different copies of the same textual version will vary slightly in word count and overall length based on the different editorial practices applied to the text.

investigation (i.e., examining salient lexical patterns within *PP*). Predominantly web-based or available via CDs, these larger pre-existing corpora function as electronic depositories of written discourse containing specific medieval text-types as seen in the *Corpus of Middle English Prose and Verse* (*CMEPV*) and the *Corpus of Early English Medical Writing* (*CEEME*), or operate as more general language corpora containing a greater sampling of linguistic content across text-types as well as chronological period represented in databases like the *Oxford Text Archive* (*OTA*). My study uses some of these corpora but only insofar as to download source-texts for the creation of my specialised poetic target and reference corpora (for more, see 3.1.2). My approach additionally utilises software and tools developed in corpus linguistics designed to extract specific linguistic features and patterns from text in ways not currently accessible via the analytical parameters of existing electronic depositories of medieval literary text (see 2.3.3 and 2.3.4).

In my computational examination of *PP* I follow general scholarship taking a combined LL and corpus-based approach (Mahlberg 2007, Norgaard et al. 2010, Ho 2011, Mahlberg & McInyre 2011, Mahlberg 2013, Mahlberg 2014, Mahlberg 2016, and Mastropierro 2017) in identifying this intersection of corpus tools, theories, and methods with literary style as *corpus stylistics*. This referent aligns with the type of analysis, investigatory aims, and general approaches described by O'Halloran (2007a) as *corpus-assisted literary evaluation,* with Fialho and Zyngier's (2014) *qualitative methodological approaches to stylistics* merging computational approaches to stylistics and stylometry under the same overarching term.[25]

[25] For the purposes of the present study, a stylometric approach using corpus linguistic methods (eg. Craig & Kinney 2009) would be referred to as *corpus stylometry*.

The following section (2.3.1) addresses the primary operationalising factors facilitated by computational approaches to style as they directly relate to the language contained in *PP*. It then further outlines the present research being conducted in (historical) corpus stylistics to date, as well as the niche the current investigation of *PP* occupies in relation to these existing studies (see 2.3.2). The remaining sections (2.3.3 and 2.3.4) introduce the two corpus linguistic approaches being used in the current study, keyword and collocational analysis.

2.3.1 Corpus stylistics

Corpus stylistics (hereafter, CS) operationalises innovations in corpus linguistics for the benefit of literary stylistic analysis. The benefits associated with using corpora to examine style are not limited to computational methodologies but crucially extend to the tools and theoretical linguistic frameworks developed by corpus approaches to language use (Mahlberg 2013). How corpus linguistics operationalises LL analysis is well-documented in existing corpus stylistic scholarship (eg. Mahlberg 2014, 2016), as well as through growing recognition of its benefits as manifested in the 'corpus turn' observable in contemporary stylistics (Leech & Short 2007: 286).

One of the primary advantages of a corpus approach to *PP* is the ability to combine qualitative and quantitative analysis of the text. The capacity of corpus tools to generate empirical data about the language use in *PP* provides an analytical counterpoint for qualitative observations in a mixed-methods approach enabling a more comprehensive investigation than possible through manual analysis alone (Carter 2014: 86, Mahlberg 2014: 387). A corpus approach to the poem builds on but does not replace existing methods of scholarly interpretation, representing 'not so much an

entirely new method as a widening of the scope of an existing one' (Mair 2006: 23), with the focus on language patterns and the effects stemming from said use mirroring much of the language-based research already visible in *PP*'s critical literary scholarship.

In addition to the efficiency of accessing fast-actioned data on the lexis within (and its distribution across) the poem, a corpus approach also provides access to new theories and descriptive categories of language use in the text. Innovations originating in corpus linguistics, these theoretical frameworks offer new types of language analysis for stylistic exploration in literary texts including the examination of keywords (Scott & Tribble 2006), lexical items (Sinclair 2004), lexical bundles (Biber et al 1999: 989, Biber et.al 2003), and narrative suspensions (Mahlberg & Smith 2012, Mahlberg et al. 2013).[26] The quantitative parameters defining said concepts make these features analytically unique to existing *PP* scholarship addressing similar types of language use on a non-empirical basis. The analytical contributions resulting from a CL approach thus offset potential criticisms of corpus applications as being merely a faster method of obtaining information that is otherwise accessible manually.

To this end, a corpus approach theoretically and methodologically operationalises the stylistic examination of *PP* on a comparative basis. This computational comparison facilitates examination using relatively large amounts of data drawn from the late 14th and early 15th century language appearing in the study's poetic reference corpus (see 3.1). The 890,778 words contained in the reference corpus does not constitute a relatively large corpus from a CL perspective, but far exceeds the scope generally utilised in the analysis of individual poetic texts. Further criticism directed at the relatively small size of fictional corpora (to which both corpora

[26] This is not meant as a comprehensive listing of theories stemming from CL but rather a sampling signposting the types of qualitative contributions a corpus approach can offer LL analysis.

used in this study belong) is contested by Lee (2010: 114), who points out that this difference in corpus size is not necessarily problematic for studies focusing on specialised language use, with 'the greater homogeneity among texts in a specialized area confer[ring] the advantage of fewer texts being required for the corpus to be representative of a language variety'.

It is this ability for corpora to represent a variety of language (McEnery et al. 2006: 4) that makes a corpus approach appealing for the present investigation. Corpus methods operate on a statistical basis, providing quantitative data about the presence (or absence) of lexical features within a text or set of texts (McEnery et.al 2006: 52). This information is derived from the automated computation of inferential statistics measuring the comparative range, variance, and deviation of the language present in the corpora (2006: 53-4). When assessed collectively, the information produced by these statistical measures provides enough linguistic information to present a reconstructed standard of language use – or language norm – representative of a historical language variety (Busse 2014: 107). While this representation of a language norm is a more relative than absolute (Mahlberg 2013: 10), these statistically-based language norms address the subjectivity inherently involved in the determination of 'conventional' language (Wales 2014a: 292) and provides a standard of late 14th/early 15th century poetic language use whose parameters are both clearly established and replicable, or 'falsifiable', by other scholars (McEnery & Hardie 2012: 14-16). This approach also offers a practical method of overcoming the challenges associated with determining normative language across the historical distance separating the late medieval discourse world of *PP*'s composition from the present day audiences (see Fitzmaurice & Taavitsainen 2007: 21-24, Fitzmaurice 2010: 699).

In providing a set of normative language against which to compare the language in *PP*, a corpus approach ultimately provides a point of departure for investigations of literary style. The normative language represented by the available corpus data operates as a linguistic standard of conventionality against which the language within *PP* can be qualitatively measured 'to establish the… stylistic realization of a particular linguistic feature or text, or to establish what constitutes creative language use generally' (Busse 2010: 37). In other words, a corpus approach enables analytical focus to be narrowed from all the language in *PP* to a selection of linguistic features identified as being unusually (in)frequent, and hence more likely to be representative of linguistic novelty or creativity. This linguistic creativity is visible in the observation of corpus-generated lexical patterns revealing irregularities (comparatively infrequent features) or extra regularities (unusually frequent recurrence of features) in the text (Leech 1969), with some of these patterns being stylistic rather than register-based.[27]

Analysing the stylistic import associated with the language used in *PP* compared to the reference corpus is assisted by corpus linguistic tools such as *AntConc* (Lawrence 2018) and *WordSmith Tools* (Scott 2012). These programmes draw on the computer's ability to actively search within the text to extract examples of specific lexical features from large amounts of text for closer analysis. This computational ability to search, retrieve, calculate, and sort through the patterns appearing within the text (McEnery & Wilson 2001: 17) provides a range of techniques for investigating specific textual features, and offers an innovative method of studying language in literature (Stubbs 2005: 22). I follow other corpus stylisticians such as Mastropierro

[27] While not using literary source texts, Studer's (2014) study of early English newspapers using a CS approach provides a useful example of how stylistic language patterns can be extracted from computationally-generated datasets for meaningful analysis.

(2017) in acknowledging that literary stylistic analysis cannot end with these computationally-generated linguistic datasets but rather needs to be the starting point for further qualitative analysis of said corpus data:

> Corpus stylistics necessitates detailed qualitative inspection to complement a quantitative breakdown, as the stylistic relevance of quantitative data cannot be assessed on the basis of statistical significance alone (Mastropierro 2017: 20).

This computationally-facilitated point of departure is particularly helpful in stylistic analysis of lengthy literary texts (Mahlberg & Smith 2010: 449). Short (1996: 255) acknowledges the novel to be one of the most difficult modern text-types to analyse stylistically, with its length making detailed analysis in its entirety more problematic than in shorter extracts like poems:

> […] the sheer bulk of prose writing is intimidating; […] In prose, the problem of how to select — what sample passages, what features to study — is more acute, and the incompleteness of even the most detailed analysis more apparent. (Mahlberg & McIntyre 2011: 206, citing Leech & Short 2007: 2)

While novels as extended prose narratives did not exist within a late Middle English literary landscape, its narrative poems were at times just as lengthy (and in some cases, lengthier) than many modern English novels. From this perspective *PP* shares more in common with novels than contemporary poetry in the difficulties presented in deciding where (and how) to start a LL analysis.

Corpus-based starting points for CS analysis tend to involve the application of LL theories and frameworks to decontextualised lists of lexical features to derive meaning from the generated corpus results. Existing CS scholarship (eg. Mahlberg 2013: 8-11) has already outlined how LL concepts such as repetition (Short 1996: 15), deviation (Leech 1985: 49), and the foregrounding effects resulting from these

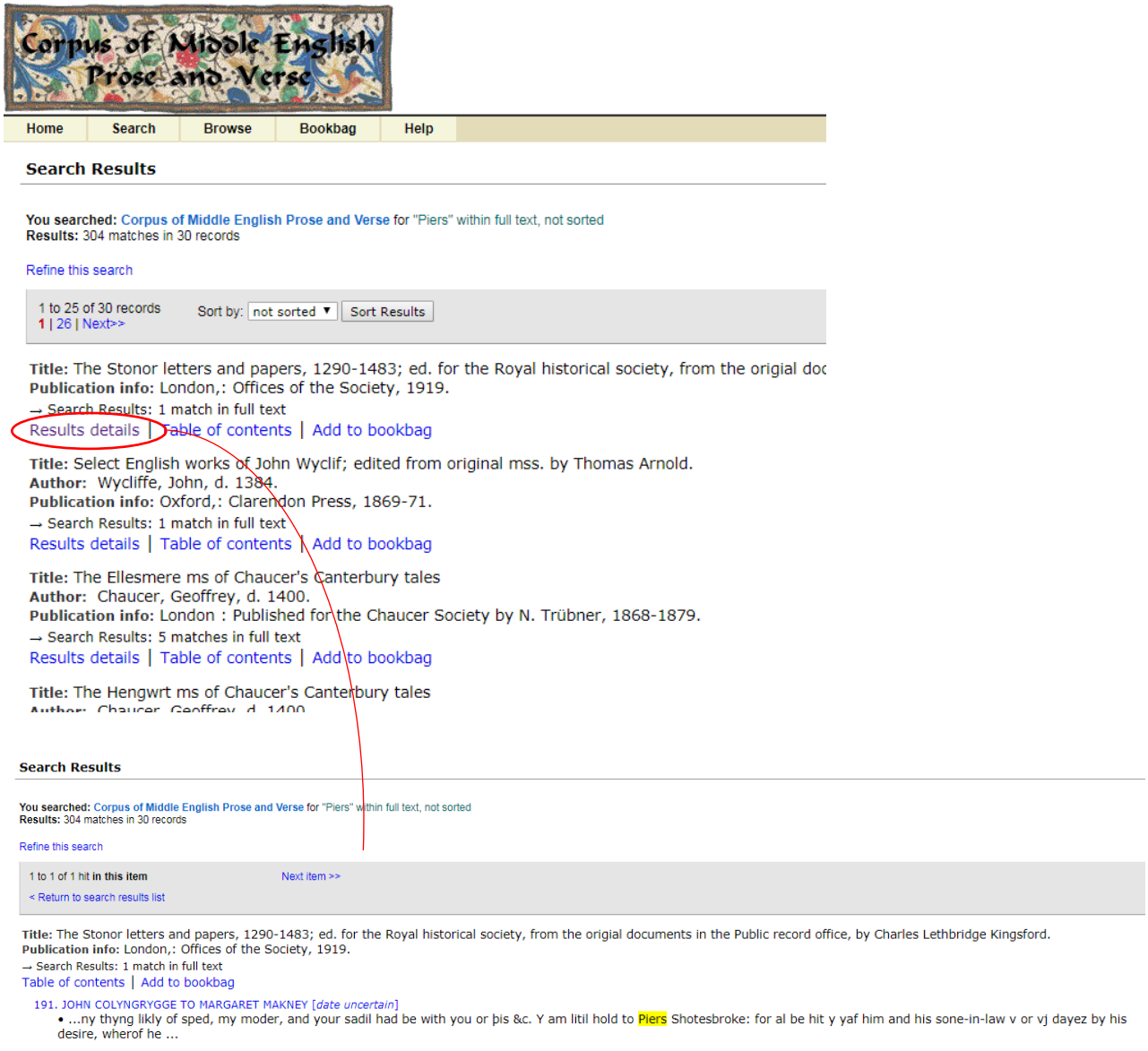
patterns (Simpson 2004: 50), among others, are extractable from corpus-derived features such as keywords (eg. Scott & Tribble 2006: 54-9, Mahlberg & McIntyre 2011: 206-7) and lexical bundles (Mahlberg 2013: 48-61). This connection of CL features to LL frameworks ascribes a corpus component to depictions of style as being something visible 'in relation to norms and deviations from norms of language use' (Mahlberg 2013: 8). The decontextualised lists containing the lexical features from which said stylistic patterns are constructed prompt 'vertical' readings of literary texts, which represent a substantively different engagement than the 'horizontal' interactions characteristic of traditional reading experiences (Mahlberg 2013: 7, citing Tognini-Bonelli 2001: 3).

It is worth briefly noting that this decontextualised approach characterising corpus-based methods is similar to traditional critical approaches to language study evidenced within medieval literary studies. These critical literary research methodologies frequently make use of the decontextualised material within concordances including, for instance, Wittig's (2001) printed concordance of *PP*. Within this context, the term *concordance* refers to an alphabetical list of the words present in a text or group of texts, with accompanying information about the originating page numbers (and where relevant, source-texts) for each term generally provided.[28] This understanding differs from a *concordance* from a linguistic perspective as all the occurrences of a search term or search phrase 'presented within the context in which they occur – usually a few words to the left or right of the search term' (Baker et al 2006: 42-4). The presented layout of these immediate linguistic contexts are moreover able to be manipulated to focus on specific patterns within the provided

[28] This type of concordance essentially acts as a type of word list that is alphabetically ordered but still contains some information of raw frequency of occurrence for each listed word (see Scott & Tribble 2006: 13-20 for discussion of alphabetically- versus frequency-ordered word lists).

window (for example, rearranging the concordance lines alphabetically not only according to the search term, but also alphabetically according to words appearing in various positions to either the right or left of the search term). While medieval scholars do have access to web-embedded concordances as defined in a linguistic sense, citational information is generally prioritised over the searchability (and comparability) of the term's linguistic context, as illustrated below in Figure 2.4.

Figure 2.4: Visualisation highlighting the citation-oriented focus of existing medieval web-based literary corpora and the attached concordances typically used by critical scholars, as represented in a search for 'Piers' within the *Corpus of Middle English Poetry and Verse*.



However, despite the advantages availed in a corpus approach to style, 'the potential of historical corpora for an explicit historical stylistic investigation has only rather tentatively been explored' (Busse 2010: 33). This observation is not limited to historical analyses but rather representative of literary linguistics more generally as a field (Mahlberg 2007). Within a modern literary context, issues of copyright can obstruct the creation of corpora containing recently published literary works, which in some cases shift analysis to corpora containing web-mined reviews of the targeted literature for a more reader response-oriented approach. Considerations from a historical dimension, by contrast, involve the availability of digital versions of the texts being studied, as well as the suitability of these electronic copies for computational analysis (see 2.4). Given the leading criticism of N-HS studies that researchers fail to adequately enegage with either ongoing LL theories or the text's literary criticism (eg. Busse 2010, 2014) the present investigation proactively includes both in the analysis chapters (5 and 6). I first use relevant linguistic theories and frameworks to analyse the generated corpus data, then in the latter part of each chapter situate these corpus findings with the literary scholarship surrounding *PP*.

2.3.2 Taking a CS approach to Piers Plowman

Existing CS analyses of historical literature follow certain predetermined patterns to which the present study partially adheres. Firstly, these examinations tend to cluster around prominent literary figures such as Austen (Fischer-Starke 2010, Mahlberg & Smith 2010, Gonzalez-Diaz 2014), Dickens (Mahlberg 2010, Mahlberg & Smith 2012, Mahlberg 2013, Stockwell & Mahlberg 2015, Ruano San Segundo 2016), and

Shakespeare (e.g. Culpeper 2009, Archer & Bousfield 2010, Anderson & Crossley 2011, Demmen 2012, Murphy 2015). Secondly, these choices are generally characterised by the prolificness of said authors, a factor useful in creating corpora of a single author's output to compare to either a) a single text written by said author, or b) a broader reference corpus of contemporary literature (if not both). Thirdly – perhaps as an indirect result of the first two criteria – CS studies tend to focus on canonical texts and major English literary periods that a modern reader would have foreseeably been expected to come across within a further or higher educational context. This contextualisation within established (and textually sizeable) literary traditions enables corpora to be constructed alongside Leech's (1985: 49) levels of deviation to reveal stylistic differences in accordance with the language norms present in a single text, a single-author corpus, or general language use.

Table 2.5: Levels of deviation set out by Leech (1985), shown with example CS analyses.

Level of deviation	**'Compare textual example to…'**	**Example CS analysis**
Primary	general language corpus	Stubbs' (2005) comparison to *Heart of Darkness* to the *British National Corpus* (*BNC*)
Secondary	all works by that author	Fischer-Starke's (2010) comparison of *Northanger Abbey* to a corpus of Austen novels
Tertiary	to whole text	Culpeper's (2002) analysis of dialogue of main characters in *Romeo and Juliet* to entire play

My approach of using a single-text, single author target corpus (the text of *PP*) and a multi-author reference corpus (a sampling of other late medieval poetic texts) most closely follows studies such as Fischer-Starke (2010), Mahlberg (eg. 2010, 2013), and Demmen (2012) in exploring secondary deviation, albeit with some key

differences. All three researchers focus on prolific authors (Austen, Dickens, and Shakespeare, respectively), which facilitates corpus comparison of a text not only via all the works by that author, but also against a broader literary corpus representative – or at least indicative – of the conventional language practices operating within a set literary tradition and time period (19th century prose for Dickens and Austen, 16th century plays for Shakespeare). As the only text thought to have been authored by William Langland *PP* cannot be studied in the same manner, leaving its reference corpus to be comprised instead of other contemporary late medieval poetic texts authored by other individuals, whether known or anonymous (for more detail on source-texts, see Tables 3.4 and 3.5).

The current investigation also follows the dominant trend in historical corpus stylistics of taking a *pattern-driven* approach (Kopaczyk & Tyrkko 2018). A pattern-driven approach operates based on Tognini-Bonelli's (2001: 10-11) *corpus-driven* approach, wherein theories or theoretical frameworks 'can only be formulated in the presence of corpus evidence and is fully accountable to it' as opposed to using corpora as 'a source of repository examples' for (or quantification of) a pre-existing theory in *corpus-based* methods. However, a pattern-driven approach departs from the theoretical neutrality expected of a corpus-driven approach to, while relying on frequency-driven corpus data, 'introduce categorisations, models and thresholds that help the researcher identify emerging patterns of language use that shed new light on the linguistic question of interest' (Kopaczyk & Tyrkko 2018: vii). This departure is important for N-HS analysis, for to take a theoretically neutral stance would be counterintuitive to modifying existing LL theories for historical literary analysis. Such a stance also represents an artificial relationship between text and researcher presupposing a lack of knowledge regarding the text's content and its greater literary

and linguistic contexts, which is not a possibility when examining historical language use without running the risk of analytical errors including overgeneralisation or overinterpretation of the corpus data.

Out of the range of available corpus methods, this project examines the language of *PP* via two corpus linguistic features, keywords (Culpeper & Demmen 2015) and collocations (Xiao 2015). The following section (2.3.3) outlines how the reasons for their selection out of the total repertoire of available corpus tools follows the preponderance of evidence in CL and CS suggesting their usefulness in revealing linguistic features likely to be of interest in a literary linguistic analysis. The current focus on literary texts necessarily omits extended discussion of keyword and collocational analyses conducted on non-literary text-types unless relevant to the present study in terms of their methodological approaches, findings, or theoretical implications.

2.3.3 Keywords

Within a corpus linguistic context, keywords are defined as terms that are observed with greater statistical (in)frequency in a text (or corpus) than would be expected by chance (Baker et al. 2006: 97) when compared to an 'appropriate sample' containing a similar language variety (Scott & Tribble 2006: 58). These terms are characterised by recurring patterns of use that 'differ significantly from their frequencies in a [given language] norm' or the patterns exhibited in another linguistic context (Culpeper & Demmen 2015: 92). In this capacity keywords are 'key' if they are significantly more frequent in the target corpus than the reference corpus (Mahlberg 2013: 63), or in their absence where otherwise expected.

Keywords (hereafter, KWs) are calculated using corpus software such as *WordSmith's KeyWords* function (version 6.0) to conduct these statistical comparisons between corpora. These programmes facilitate comparison by making it

> a relatively easy and rapid task for a researcher to calculate the incidences of each and every single word in the target data as well as a comparative dataset, undertake statistical comparison between incidences of the same words in order to establish significant differences, and see the resulting keywords ranked according to degrees of significance of difference (Culpeper & Demmen 2015: 93).

The decontextualised KWs are vertically represented in lists displayed in descending order of generated keyness, with terms having the highest degree of unusual recurrence (in this case, within *PP*) located at the top of these keyword lists. These keyness values indicate the strength of each KW's unusual recurrence (or in the case of negative KWs, absence) in the text. In some of these KWs, these measures provide a statistical basis for qualitative understandings of *keyness* as 'a quality words may have in a given text or set of texts, suggesting that they are important… reflect[ing] what the text is really about, avoiding trivia and insignificant detail' and revealing 'what a text boils down to' (Scott & Tribble 2006: 54-5).[29] While several different statistical tests are available for measuring keyness, the resulting KW lists appear to be more affected by the composition of the reference corpus than the statistical test itself, 'with only minor and occasional differences in the rankings of keywords, differences which had no effect on the overall picture revealed by the keywords' (Culpeper 2009: 36). KWs can also be visualised in dispersion plots indicating the distribution of the term's use within a given corpus (or in the case of a single-text corpus like *PPCorp*, across an individual literary text).

[29] In this sense, a corpus approach to keywords provides an empirical basis for comparison to culturally-derived keywords (eg. Williams 1983) with similarities in purpose of identifying 'key' terms, if not in the procdures involved in their identification (Rayson 2008: 523, citing Scott 1997).

The benefit of KW analysis lies in the ability to identify lexis that is distinctive to, or characteristic of, a single text in relation to other texts (Baker 2006: 125, Stubbs 2005: 11, Culpeper & Demmen 2015: 99). KWs point towards tendencies of lexical usage that are unique to *PP*, which are different to measures of frequency; rather than counting the raw occurrence of a term in *PP* in isolation from its greater literary context, a KW approach 'compares each word type within the text in question to the same word type in the reference corpus to reveal instances of unusual lexical recurrence that occurs to a degree that is statistically significant' (Scott 2001: 155). This is not to claim that frequency information does not reveal interesting insights about literary texts, but rather that a KW approach offers a 'more sensitive' (McCarthy & Handford 2004: 174, cited in Harvey 2013: 58) quantitative measure indicating the *degree* of importance attributable to terms within literary text relative to their comparative presence (or absence) in a broader discoursal context.

In this capacity, KWs constitute a simplified way of examining meaning in literary texts (Mahlberg & McIntyre 2011: 208), including *PP*. The lexis-driven approach facilitated by KW analysis has the potential to gain access to aspects of the text that are integral but either obscured or otherwise not intuitively obvious (Bondi 2010: 5). KW approaches are also useful on a practical basis when approaching lengthy texts like *PP*, whose size makes identifying an initial lexical clustering of characteristic terminology challenging without computational assistance reducing focus to a smaller and more concentrated language sample (Mahlberg 2009: 50). Generating KWs for the poem perforce narrows the scope of analytical focus from all the words contained in the text (70,541) to a shorter list of lexis (855) identified as being fundamentally characteristic to (and within) *PP* on a quantitative basis.

KW analysis is appealing to LL investigations based on the ability to extract lexical markers of style from generated KW lists. Corpus linguistic scholarship (eg. Scott & Tribble 2006: 54, Mahlberg 2013: 67) broadly identify three categories of KWs: proper nouns, aboutness indicators, and style indicators. These categories signpost the ability of KWs to highlight both semantic content and stylistic patterns within a literary text on a lexical basis, with aboutness indicators reflecting the content or the general 'aboutness' of the text compared to stylistic markers signalling how said content is conveyed. Culpeper (2009: 39) points out one of the main issues posed in dualist approaches treating stylistic markers as fundamentally separate from content (i.e., aboutness markers as lexical words and style markers as grammatical terms), which can potentially result in the inadvertent marginalisation of style as mere 'decoration'. He argues that a functional Hallidayan approach labelling KWs as being ideational, textual, or interpersonal (eg. Halliday 1985 [1994], Halliday & Matthiessen 2004) captures the 'functional essence' of each KW and views 'all choices, including grammatical choices, as meaningful and stylistic' (2009: 39). Culpeper (2009) demonstrates how this approach is effective by examining key features like interjections and vocatives associated with specific characters in Shakespeare's *Romeo and Juliet* (in the KWs generated for Nurse in Culpeper 2002, for instance) which are important to characterisation within the play but are difficult to meaningfully categorise as being exclusively aboutness or stylistic markers.

This access to lexical stylistic markers has resulted in a degree of observable popularity attributable to KW analysis within existing CS investigations (for a general survey, see Demmen 2012: 40-2, Mahlberg 2013: 20-1). While there is presently no precedent set for using keyness to assist in analysing lexical style in Middle English poetry, the findings of other historical CS analyses suggest a large scope of analytical

potential in its application. Fischer-Starke (2010), for example, uses KWs to identify lexical patterns in *Northanger Abbey* compared to both a corpus of Jane Austen novels, as well as a corpus containing contemporary 19th century prose fiction. In identifying a series of key terms that are significant to the novel Fischer-Starke (2010: 26) demonstrates that characterisation of the story's protagonists is linked to their reading habits, 'an observation new to the criticism of the novel, because it can really only be readily observed with this kind of technologically-supported analysis'. This use of keywords to facilitate analysis of characterisation is also evident in studies including but not limited to Culpeper (2001, 2002, 2009), O'Halloran (2007b), McIntyre (2010), and Walker (2010).

In another CS study Mahlberg and McIntyre (2011) survey corpus-based analysis of keyness in literary texts, including KW analysis. While exploring modern popular fiction (eg. Ian Fleming's *Casino Royale*) over historical text-types, the study reveals the types of stylistic study that can be conducted on literary texts through a classification system for lexical functionality separating KWs into 'fictional world' and 'thematic' functions. Fictional world terms constitute the concrete words that develop a text's fictional word, a concept not unlike the world-building elements in text-world theory (Gavins, 2007: 36-8). These terms establish the text world's setting physically, temporally, and spatially, and additionally convey information about the fictional characters appearing (and events taking place) in said fictional world. Thematic words, by contrast, are less concrete in that they retain additional thematic significance to a text beyond their core denotational meaning. Mahlberg and McIntyre (2011: 209) point towards *gambler* as a thematic term that, in a James Bond novel, ostensibly exists as part of an extended theme in the text beyond identifying individuals characterised by participation in gambling-related activity. The creation of this

theoretical framework integrates LL with corpus methods to achieve ways of obtaining new insights into literary texts which is fundamentally transferrable across literary text-types, an approach since extended to other modern (eg. Yazdanjoo et al. 2016, Jaafar 2017) and historical novels (eg. Mastropierro 2017) as well as Middle English poetic texts (Cordell 2018).

Taking a KW approach to *PP* answers the current research's aim to identify lexical features distinctive to the poem (RQ1 in 1.2), and provides a point of departure for further analysis of literary style through KW concordances. As briefly discussed in 2.3.1 concordances reveal the immediate linguistic environment for every occurrence of a keyword within a given text or corpus to provide a 'keyword in context' (KWIC) view. This limited contextualisation enables qualitative examination of quantitatively measured features through the 'associated text extracts' in which these keywords occur (Mahlberg & McIntyre 2011: 206) in a view 'allowing linguistic patterns to be more easily observed' (Baker et al. 2006: 43), as exemplified below in Figure 2.6.

Figure 2.6: Concordance of 22 out of the 42 total occurrences of the word *amend* in *Piers Plowman*, listed alphabetically in descending order according to the word appearing in the L1 position.

medicine most thereto that might	**amend**	the prelates That should pray f
id he no man as y leave He might	**amend**	in amynte-while al that amys st
is the best Until that mischief	**amend**	hem that many man chastise For
kind sesed tho to se the people	**amend**	Fortune began flateren then to
e and to thine in hope you shost	**amend**	And always the lengur y late th
nd prayed thus the people should	**amend**	That contrayen contraryen now c
nd resceyued for the worthier To	**amend**	and to make as with men of holy
here piteously to pray for me to	**amend**	And also ken me kyndly on chris
Have this man of me said she to	**amend**	thy sh harm For y will wage for
that he nor is full of might To	**amend**	al that amys is and his mercy g
die Quietly for our misdeeds to	**amend**	vs all And ?ut would hem no wo
ive hem grace here doing here to	**amend**	And thus is dowel my friend to
d bicam man of a maid mankind to	**amend**	And made yourself with thy son
Grant gylours on earth grace to	**amend**	And have here penance on pure e
ynnefole our saviour of grace To	**Amend**	vs of our misdeeds do mercy to
nsufficient faith and no will to	**amend**	For a leaves be lost when he hi
r his goodness give the grace to	**amend**	Said repentance right with that
ad him bid to god be his help to	**amend**	Then said lechery alas and to o
okene tobrokene by the high ways	**Amend**	in some manner wise and mayndon
ting in bad men but if they will	**amend**	dread not for no death to destr
Both wyndowes and immense y will	**amend**	and glaze And peynten and purtr
essing do vs both but if we will	**amend**	Nor mans prayer makes pees amon

Mahlberg and Smith (2010: 466) outline the benefits of concordances within corpus stylistics in 'retriev[ing] data that is relevant to the question [guiding] analysis' but stress the need for the analyst to evaluate the relevance of the provided data and 'to assess how searches can be narrowed down and further complemented with further methodological approaches' (see also Rayson 2008: 528-9).

Identifying KWs for further concordance analysis is achievable by categorising the poem's key terms into semantic domains. Semantic domains extend analysis from unusually frequent words to unusually prominent topical domains containing subsets of key terms sharing similar meaning and content (Mahlberg 2013: 20). These semantic domains help corpus studies identify and compare major trends within discourses of

language use, in addition to stylistic comparisons facilitated at the word level (Rayson 2008: 542).

Within stylistic analysis, semantic domains have predominantely been generated computationally using semantic taggers. For example, Culpeper (2009), Archer and Bousfield (2010), and McIntyre (2010) use the corpus tool *WMatrix* (Rayson 2009) to extend studies of literary keyness from KWs to key semantic domains in Shakespeare's *Romeo and Juliet* (Culpeper 2009) and *King Lear* (Archer & Bousfield 2010), as well as Quentin Tarantino's screenplay *Reservoir Dogs* (McIntyre 2010). In this approach KW categorisation occurs through a process wherein corpora are semantically annotated for automated separation into major discourse fields – and sub-divisions within each field – according to universal classification guidelines like the *UCREL Semantic Annotation System* (*USAS*) (Rayson 2009). Given the modern focus of the language categories in *USAS*, historical key semantic domain analysis has become more viable through development of the SAMUELS Project's *Historical Thesaurus Semantic Tagger* (*HTST*) as a sense disambiguator for thematic categories reflecting different socio-cultural dimensions of meaning within more historical contexts of language use (see Piao et al. 2014, Rayson et al. 2015).

The present investigation follows existing CS studies in using semantic domains to establish key thematic topics within a literary text within which to situate a language-based stylistic analysis. Extending analysis of keyness beyond individual lexical units within *PP* is also beneficial in engaging with the critical literary scholarship surrounding the text, as much of it is already organically oriented around semantic- and thematic-based categories (for example, themes of law, government, poverty, labour, or consumption). This methodological alignment with *PP*'s surrounding criticism thus facilitates analytical discussion within an interface between linguistic

form and functional import, an associative approach moreover continued in the decision to utilise relevant semantic frameworks drawn from the literary criticism for KW semantic categorisation rather than basing analysis on automated methods of corpus-based data retrieval like *WMatrix*. While losing the objectiveness associated with automated approaches, this decision facilitates more direct comparison of the present study's findings to existing discussion of the poem and its language in relevant Langlandian scholarship.

2.3.4 Collocations

As with keywords collocation represents another statistically-based feature of language use, in this case measuring the association between lexical units. Collocates are words that correspond with one another based on their mutual occurrence in similar linguistic environments (Firth 1968: 181-2) which form collocations, or 'the characteristic co-occurrence patterns of words' (McEnery & Wilson 2001: 85). These collocations are computationally-retrievable text features consisting of *nodes* (the targeted search term being studied), and the contextual *spans* in which the node's collocates are located:

> We may use the term node to refer to an item whose collocations we are studying, and we may then define a span as the number of lexical items on either side of a node that we consider relevant to that node. Items in the environment set by the span we will call collocates (Xiao 2015: 107, citing Sinclair 1966: 415).

The current study focuses on the idea of *neighborhood collocation* (Sinclair 1996). These collocates do not have to be textually consecutive to the given node but can extend to word associations operating within a set span, or window, of textual

proximity. While collocational relationships between a node and its collocates are in some cases intuitively recognisable (e.g., the collocates *pickings* and *chance* to the node word *slim*), 'it is more reliable to measure [collocation] statistically, and for this a corpus is essential' (Hunston 2002: 68). This statistical basis is particularly important for historical varieties of language use where qualitative identification of typical collocational patterns, like with language norms, presents challenges to detailed language-based analysis.

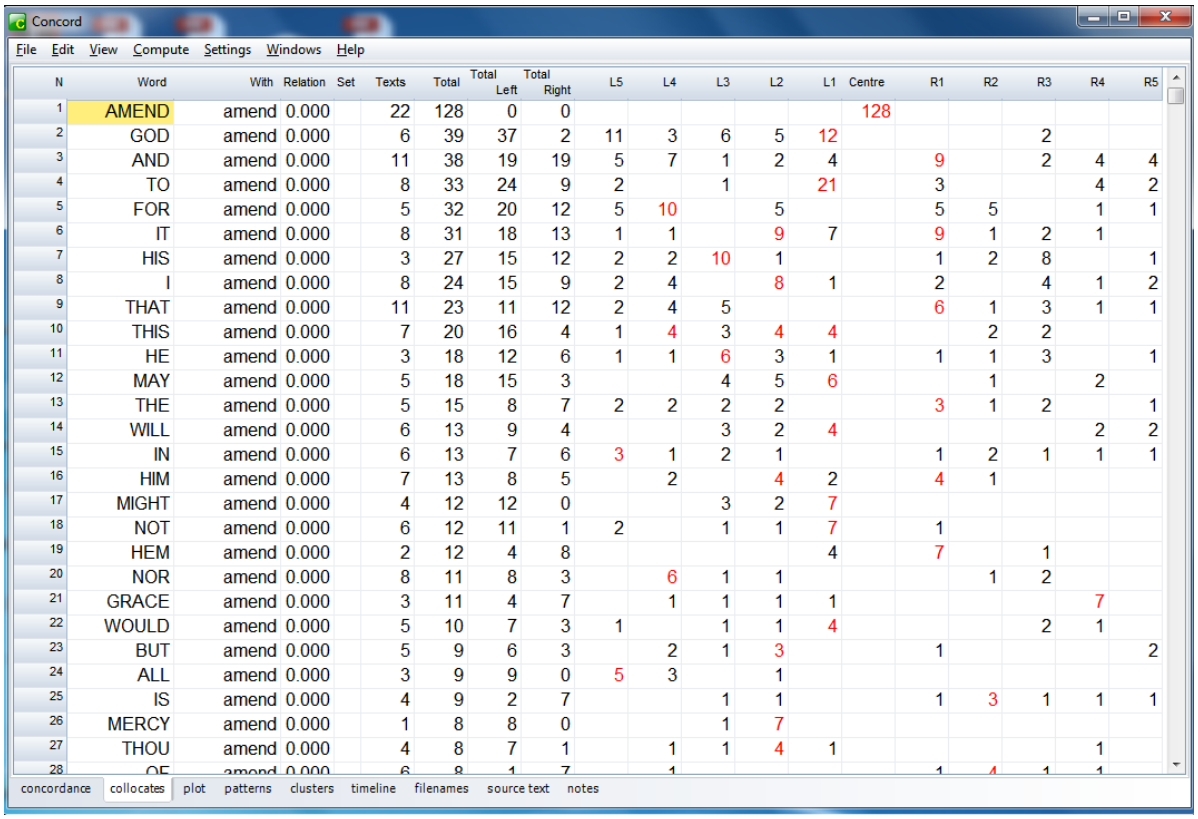
Collocates are also generated using *WordSmith* (version 6.0), which contains built-in functions offering the analyst a variety of statistical measures by which to evaluate collocation. Each statistical test reveals different information about lexical proximity: t-test measures reflect the confidence by which a collocational relationship exists, while mutual information (MI) scores reveal the strength of said collocations (Church & Hanks 1990). Xiao (2015: 110) observes that log-likelihood (LL) produces the best results for analysis with smaller amounts of text, which suggests the suitability of that measure for the typically small specialised corpora used in literary stylistic analysis. Using statistical texts like LL or MI operate as part of the 'collocation-via-significance' approach, which stands in contrast to the 'collocation-via-concordance' approach characterised by using raw frequency accounts of co-occurrence (McEnery & Hardie (2012: 126-7). Collocation-via-significance is generally considered an analytically stronger approach, but the smaller size of some specialised corpora sometimes result in limited numbers of collocational recurrence, which requires the researcher to use a collocation-via-concordance approach instead.[30]

Like with KWs, generated collocates are provided in decontextualised lists arranged in descending order of ascertained textual affinity. These listings can be

[30] For an example, see Mastropierro (2017: 114-5).

extended into a view illustrating the textual positions in which these collocates are situated relative to the node within a set span. Concordances can also be used to view the longer text extracts containing these identified collocational patterns in instances where further contextualisation is required.

Figure 2.7: View of the collocational patterns for *amend* in the poetic literary reference corpus within a 5-word span to the left and right of the node word, as generated in *WordSmith Tools.*


Concord

File Edit View Compute Settings Windows Help

N	Word	With	Relation	Set	Texts	Total	Total Left	Total Right	L5	L4	L3	L2	L1	Centre	R1	R2	R3	R4	R5
1	AMEND	amend	0.000		22	128	0	0						128					
2	GOD	amend	0.000		6	39	37	2	11	3	6	5	12				2		
3	AND	amend	0.000		11	38	19	19	5	7	1	2	4		9		2	4	4
4	TO	amend	0.000		8	33	24	9	2		1		21		3			4	2
5	FOR	amend	0.000		5	32	20	12	5	10		5			5	5		1	1
6	IT	amend	0.000		8	31	18	13	1	1		9	7		9	1	2	1	
7	HIS	amend	0.000		3	27	15	12	2	2	10	1			1	2	8		1
8	I	amend	0.000		8	24	15	9	2	4		8	1		2		4	1	2
9	THAT	amend	0.000		11	23	11	12	2	4	5				6	1	3	1	1
10	THIS	amend	0.000		7	20	16	4	1	4	3	4	4			2	2		
11	HE	amend	0.000		3	18	12	6	1	1	6	3	1		1	1	3		1
12	MAY	amend	0.000		5	18	15	3			4	5	6			1		2	
13	THE	amend	0.000		5	15	8	7	2	2	2	2			3	1	2		1
14	WILL	amend	0.000		6	13	9	4			3	2	4					2	2
15	IN	amend	0.000		6	13	7	6	3	1	2	1			1	2	1	1	1
16	HIM	amend	0.000		7	13	8	5		2		4	2		4	1			
17	MIGHT	amend	0.000		4	12	12	0			3	2	7						
18	NOT	amend	0.000		6	12	11	1	2		1	1	7		1				
19	HEM	amend	0.000		2	12	4	8					4		7		1		
20	NOR	amend	0.000		8	11	8	3		6	1	1				1	2		
21	GRACE	amend	0.000		3	11	4	7		1	1	1	1					7	
22	WOULD	amend	0.000		5	10	7	3	1		1	1	4				2	1	
23	BUT	amend	0.000		5	9	6	3		2	1	3			1				2
24	ALL	amend	0.000		3	9	9	0	5	3		1							
25	IS	amend	0.000		4	9	2	7			1	1			1	3	1	1	1
26	MERCY	amend	0.000		1	8	8	0			1	7							
27	THOU	amend	0.000		4	8	7	1		1	1	4	1					1	
28	OF	amend	0.000		6	8	1	7		1					1	4	1	1	

concordance collocates plot patterns clusters timeline filenames source text notes

Studying collocation is useful in extending the scope of lexical analysis from individual words to recurring lexico-grammatical patterns. In general language use collocational patterns provide the necessary context to disambiguate polysemous terms and to identify close synonyms by showing not only 'patterns of lexical association' but also 'how a word can acquire meaning in context, which may differ from or even contradict its literal meaning' (Xiao 2015: 115). In this capacity

collocations reveal levels of generated meaning not discernible from single words in isolation, as exemplified in studies including but not limited to Sinclair's (1991) examination of the pragmatic significance embedded in *break out,* Louw's (1993) study of *bent on*, and Hunston's (2002) discussion of *sit through*. Collocation is thus representative of one of four hierarchical levels of extended meaning outlined by Sinclair (2004), which have been replicated below in Table 2.8.

Table 2.8: Levels of meaning generated from the relationship between lexical units, as set out by Sinclair (2004)

Collocation	the relationship between a node and individual words
Colligation	the relationship between a node and grammatical words, categories, or patterns
Semantic preference	Semantic sets of collocates
Semantic prosody	Affective meaning of a given node with its typical collocates

Within a literary linguistic context, this access to levels of pragmatic meaning – as determined by corpus-generated lexical features – enables focus on the stylistic import embedded in how collocational patterns produce semantic prosody within (and across) literary texts. Defined as the 'meaning arising from the [recurring] interaction between a given node and its typical collocates', semantic prosody is useful in literary analysis based on its primary function signposting information about speaker attitudes and evaluative stances (Louw 2000: 57-8). Stubbs (2005: 14-18) advocates for prosodic examinations of literature based on their ability to create connotations generated from collocational patterns. These, he argues, produce meaning having either pragmatic or discoursal significance, and demonstrate how prosody contributes to complex stylistic features including metaphor construction, the evaluative attitudes attributed to fictional characters, and instances of intertextuality.

Exploring *PP*'s collocations thus seeks to identify some of the key collocational patterns and associated semantic prosodies evidenced from a corpus approach to the text. This focus is 'key' in that the nodes selected for collocational analysis are derived from the KWs initially generated for the poem compared to the reference corpus, and within these, terms selected from the semantic domains chosen for further analysis (in the case of the present study, themes addressing the law and dietary consumption). Given that *PP* was composed during a time of historical political, religious, and social upheaval (2.1.2), the poem is likely to contain prosodic information regarding contentious topics addressed in plot (e.g., poverty, social repression, sinful behaviour, and salvation) whose evaluative significance can be pinpointed at the level of text through the collocational patterns within which these KWs are textually situated. Using collocations moreover addresses the limited scope of KW analysis and its focus on single lexical units to extend assessment of the poem's language to recurring multi-word sequences (Stubbs 2005: 15-18, Culpeper & Demmen 2015: 115) and their potential stylistic significance.

2.4 Studying historical language using corpora: some considerations

The lack of medieval literary stylistic analyses using corpus methods to date is based on a number of practical, theoretical, and technological factors. Firstly, analysis is reliant on the language appearing in written discourse surviving into the present day to which modern researchers have access. These texts are neither balanced nor representative of general language use for a particular time period (i.e., disproportionate amounts of high literature and official records and only limited representations of everyday discourse), but rather constitute a 'random sampling'

based on accidental or incidental survival of fires, floods, and other destructive forces (Taavitsainen et al. 2000: 1). Secondly, the amount of surviving Middle English literature is substantively smaller than the available corpus of modern literature; this can make comparative approaches like corpus analysis, which rely on large amounts of data, a challenge to justify the potential significance or extent of occurrence of a particular feature or text pattern, even when operating within highly narrowed research parameters.

In addition to the amounts and types of available historical language, historical CS studies also encounter a number of theoretical challenges. Corpus methods address the temporal divide between text and reader by providing statistically-based empirical measures from which information about historical language norms, degrees of relative keyness, and collocational significance can be extrapolated for information about language innovation in a stylistics context, but there still remains an absence of literary linguistic theoretical frameworks specifically designed for historical language. The inability to immediately confer modern LL frameworks or classification systems to historical texts (Busse 2010: 35) makes it difficult to facilitate growth in historical corpus stylistic scholarship without adaptation.

However, the most substantive intervening factors have been technological, for instance the digital availability of historical literary materials. This availability does not refer to whether the text of a manuscript can be accessed electronically but rather alludes to the existence of digital copies within machine-readable formats compatable with corpus tools and software. This formatting is crucial in enabling the computer to 'carry out the processes of searching for, retrieving, sorting, and calculating linguistic data' (McEnery & Wilson 2001: 17) within the given file(s), which is not possible in image-based versions of text like PDFs. Electronic versions of medieval literary texts

are starting to become more readily available in machine-readable formats, but focus on OE and ME manuscripts in the digital humanities still appears to be predominantly geared towards producing image-based digital facsimiles which, while interesting from a manuscript studies-based perspective and good in terms of accessibility, offers little benefit to text-mining approaches such as corpus analysis.

Existing corpus methods moreover encounter challenges when working with datasets characterised by high amounts of spelling variation like those evidenced in Middle English. This constitutes one of the biggest challenges to not only stylistic analysis but all historical linguistic spheres using corpus methods (e.g. pragmatics, sociolinguistics, dialectology). The following section (2.4.1) outlines the issues associated with corpus linguistic analysis of Middle English, and methods of overcoming the challenges its irregular orthography presents (2.4.2).

2.4.1 Spelling variation

The significant degree of spelling variation present in English texts written before 1500 constitutes one of the main obstacles to studying medieval literature through corpus methods (Markus 2002, Diemer 2012: 1-2). Variant spelling, or systematic irregularity in spelling practice, is a prominent text feature distinctive to materials composed before orthographic practices were gradually standardised in English between 1500 and 1700 (eg. Vallins & Scragg 1965, Gorlach 1991, Nevalainen 2006). An ideological as much as a linguistic phenomenon (Mair 2006: 5), spelling standardisation as a notion wherein importance is attached to individualised spellings for each word is taken for granted in contemporary written discourse but in reality constitutes a relatively recent historical development to orthographic practice in English (Horobin 2013). My corpora

of Middle English poetry predates this mentality, with most texts having been composed well before any active and sustained movements towards standardisation in the language.

The spelling variation present in my corpora is largely due to the to the phonetically-based orthography characteristic of the time period. This practice of representing words by the textual form best capturing how it would sound when spoken aloud incorporates some of the significant sound changes occurring in late Middle English including stress-related vowel changes, breaking, and initial stages of the Great Vowel Shift; these ongoing changes produced a wide variety of spellings, particularly observable in the representation of vowels and semivowel sounds (Baugh & Cable 2002: 146-55). These changes occurred during (and in doing so, contributed to) a linguistic 'interregnum' (Blake 1996: 136) wherein the former written standards governing Old English had disappeared and a new standard, in the process of being developed, had not yet established itself on a wide scale. Additional factors contributing to irregular spelling in Middle English include dialectal differences based on geographic region, the absence of overarching orthographic guidelines influencing scribal practice (and thus, scribal error),[31] and spelling decisions made for metrical reasons (e.g. the addition or removal of letters for line justification) (Scragg 1974 and Nevalainen 2006: 29-30, cited in Lehto et al. 2010: 280; see also Vallins & Scragg 1965: 71, Blake 1996: 132-171, Markus 2000: 182-3, and Baron & Rayson 2008).

While spelling variation can be of analytical significance in linguistic investigation, abnormal spelling is problematic in that it reduces the ability of corpus tools to generate accurate results based on orthographic matching (Baron et al 2009a:

[31] In this capacity I refer to scribal 'error' as the mistaken substitution of a word for a close approximant (usually identifiable by contextual parsing) rather than the inclusion of textual features indicative of scribal idiolect, where possible to separate the two.

41, Baron & Rayson 2009: 1-2). On the most fundamental level orthographic matching occurs in single-string searches for lexical units wherein the computer only retrieves results that exactly match the given search query; high amounts of irregular spelling in this context perforce prevents all total occurrences of a word within a given text or corpus from being generated from a single seach (Baron 2008). In other words, without orthographic consistency corpus tools have no basis for understanding that two variants appearing in different word forms actually represent the same lexical unit.

In this capacity spelling variation effectively prevents accurate identification of the number of times a single word occurs in a corpus, resulting in situations wherein individual words get erroneously divided into multiple lexical entries within corpus-generated word lists. Unlike keyword lists, word lists lists catalogue all lexical forms present in a text according to frequency of lexical occurrence, with frequently used words given a higher rank in the list than words appearing only once in the text (Scott & Tribble 2006: 10). The misidentification of variant spellings as unrelated word forms thus impacts the frequency measures assigned to each lexical unit by the corpus tools, as exemplified below in the representation of 'Gawain' as both the 56th and 87th most frequent word(s) contained in the poem *Sir Gawain and the Green Knight*.

Figure 2.9: Screenshot from the word list generated for *Sir Gawain and the Green Knight* showing the automated separation of the variants *Gawayn* and *Gawan* in *WordSmith.* The Freq. column lists the raw frequencies of each word, with the N column indicating the word's rank in the text based on this frequency. [32]

N	Word	Freq.
56	GAWAYN	44
57	GOD	43
58	HO	43
59	NEUER	41
60	SIR	40
61	HER	38
62	MORE	38
63	BOÞE	35
64	SYÞEN	35
65	MAY	34
66	WOLDE	34
67	ABOUTE	33
68	SAYDE	33
69	BURNE	31
70	VP	31
71	WHEN	31
72	AR	30
73	IF	30
74	SCHULDE	30
75	BY	29
76	ÞY	29
77	FRO	28
78	FYRST	28
79	BEST	27
80	MUCH	27
81	AFTER	26
82	OUT	26
83	ÞERE	26
84	VCHE	26
85	DAY	25
86	FAYRE	25
87	GAWAN	25

This division of frequency measures between variants causes a knock-on effect in conducting further analysis of corpus features like keywords and collocation (Baron 2008, Baron et al. 2009a, Baron et al. 2009b, Lehto et al. 2010). Generating keywords

[32] Data taken from my unpublished Masters dissertation (Cordell 2014)

and collocations relies on the frequency counts established in the initially generated word lists for a corpus (in this case of the present study, all the words contained in *PP*) to determine the statistical significance of language use compared to other texts (Scott & Tribble 2006: 55). If the initial raw frequencies generated in the word lists are distorted as a result of orthographic variation, then the resulting list of keywords or generated collocational relationships between lexical units risks an inaccurate portrayal of what is taking place at the textual level (Baron et.al 2009b: 47-8, 54-8). These adverse effects thus create an impact significant enough to make a corpus linguistic approach to historical language untenable in terms of data quality without somehow addressing the spelling variation embedded in historical datasets.

2.4.2 Orthographic normalisation

One method of addressing the high amounts of variation in *PP* lies in orthographic normalisation. Orthographic normalisation, or regularisation, alters lexical representation within a given linguistic dataset with the aim of reducing the number of contained variants. This reduction of variants improves the accuracy of lexical classification using corpus tools, which improves the reliablility of the frequency-based statistical measures utilised for subsequent keyword and collocational analysis. I follow Archer et.al (2015) in eschewing the term *standardise* in preference for *normalise* or *regularise* to avoid the prescriptive elements ascribed to spelling standards and standardisation that do not apply to a 14th century mentality, nor its resulting orthography; my aim in the present investigation is not to impose a set of established spelling conventions on my texts, but rather to create enough orthographic uniformity to better support a corpus analysis of Middle English poetry.

Corpus linguistic approaches to normalisation prototypically involve corpus annotation. Annotation, or *tagging*, constitutes the corpus-wide automated or manual layering of additional information to the original raw text (Baker et.al 2006: 13, McEnery & Wilson 2001: 32). In addition to making specific sets of linguistic features searchable within a corpus[33] annotation provides a method of corpus preparation, particularly in corpora demonstrating extensive spelling variation (Archer & Rayson 2006). While some corpus linguists (notably, Sinclair 2004) caution against the application of annotation to corpora,[34] I follow studies such as Culpeper and Kyto (2010) and Demmen (2012) in choosing to annotate as an important step in corpus-assisted analysis of historical corpora containing considerable amounts of irregular spelling, which Markus (1996: 222) also argues to be 'an un-avoidable pre-condition for any reliable corpus-based analysis' of pre-1700 text.

Orthographic annotation can be achieved through recently developed annotation tools such as *Variant Detector* (hereafter, *VARD*). First developed at Lancaster University by Alistair Baron (2008), *VARD* is designed to be used a) as a pre-processor prior to corpus analysis, and b) to deal with the spelling variation present in corpora, particularly within datasets of historical language (Baron & Rayson 2009: 2-6). The tool uses several linguistic algorithms (as detailed in Rayson et.al 2007: 4-10) to pinpoint spelling variation, identifying possible variants either by being absent from the programme's internal word bank, or in being approximate in lexical form or phonetic arrangement to a closely-related word. It then uses techniques drawn from modern spellchecking to offer suggestions as to the word's original spelling, with the

[33] This type of annotation is different to orthographic annotation because it identifies features which the researcher wants to examine further in an otherwise useable corpus, rather than attempting to make a dataset viable for corpus analysis. Annotation can be used to examine features including (but not limited to) parts of speech (Rayson et al. 2007), dialogue or speech acts (Culpeper 2009, 2014), and semantic domains (Rayson 2008).

[34] See also the concerns raised by Rissanen (1989).

option of manually inserting the correct spelling should it not be offered automatically (Baron et al. 2011).

The benefits of *VARD* lie in machine learning, with the capability of the tool to (accurately) recognise spelling variants and substitute them for normalised forms growing proportionally to the amount of manual training it receives from the given researcher. This type of machine learning enables annotation to move from a manual to automated process once the tool has been trained on a specific variety of language (Baron & Ryason 2009). This training allows *VARD* to conduct automated corpus-wide annotation (i.e., variant substitution) according to a pre-set statistical threshold of confidence, meaning that regularisation does not require the analyst to manually normalise each textual instance of irregular spelling (see 3.2). While possible to only use the manual annotation component of the tool, a mixed approach of using its manual and automated components combined has been recommended as best practice (Archer et al. 2015: 19).

Since its release *VARD* has been shown to improve the quality of corpus analyses involving linguistic data expected to contain orthographic inconsistency. The first wave of scholarship using *VARD* consists of a series of pilot tests conducted by corpus linguists from the school coming out of Lancaster (e.g. Archer et.al 2003, Rayson et.al 2007, Baron et al. 2009a, Baron et.al 2009b, Baron et.al 2011) which provide in-depth statistical, inferential, and settings-based comparisons of *VARD*ed versus non-*VARD*ed material. Predominantly focused on the variation observed in Early Modern English for which the tool was originally designed to address (Rayson et.al 2007: 4), these studies reveal how pre-analysis orthographic annotation impacts the quality of further corpus methods including part of speech tagging (Rayson et.al 2007), semantic tagging (Archer et.al 2003, Baron et.al 2015, Alexander et.al 2015),

and the production of keyword frequencies (Baron et.al 2009b, Baron et.al 2011). This improvement is additionally illustrated in historical corpus investigations of non-literary texts, including late 15th and early 16th century medical and legal discourse in the *Lampeter Corpus of Early Modern English Tracts* (hereafter, *Lampeter*) (Rayson et.al 2007), 16th century medical texts in the corpus of *Early Modern English Medical Texts* (hereafter, *EMEMT*) (Hiltunen & Tyrkko 2013), and 17th century medical texts in *Early English Books Online* (*EEBO*) (Ratia 2013).[35]

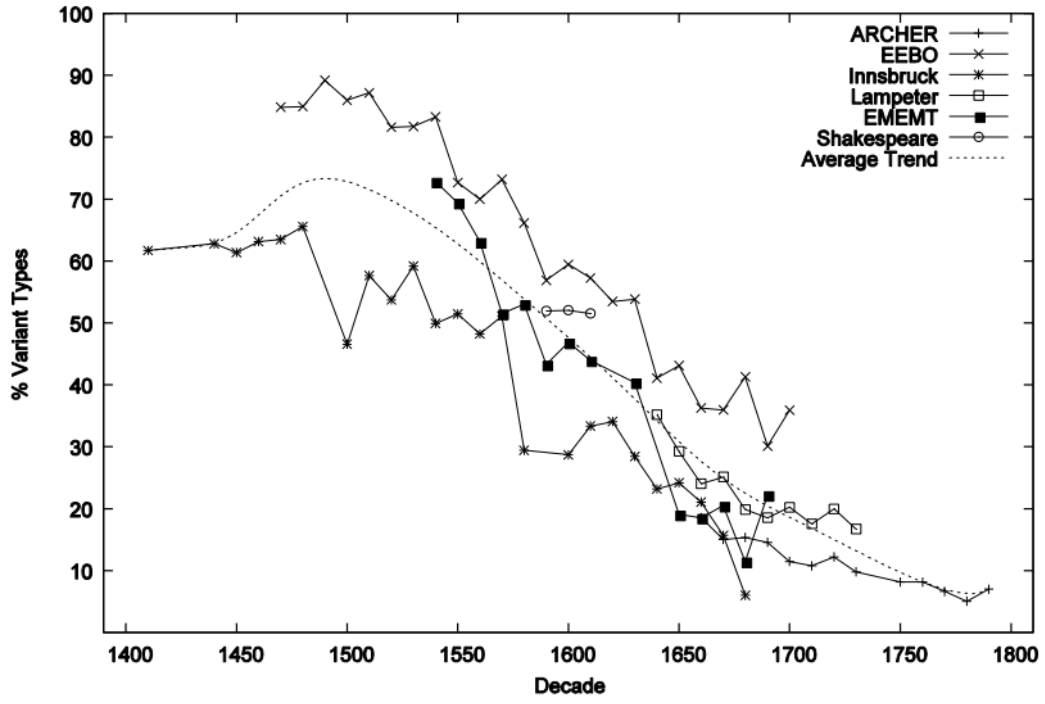
Several historical CS analyses examining early modern language and literature have approached orthographic variation using corpus annotation. To date two CS analyses incorporate *VARD* as a pre-analysis step in corpus creation: these include Culpeper and Kyto's (2010) investigation of speech representation in Early Modern plays, didactic works, witness depositions, and prose fiction (within which style constitutes one of several factors contributing to the written depiction of spoken language), and Demmen's (2012) stylistic analysis of Shakespeare's plays through keywords and lock words (Baker 2011). These investigations are representative of the second wave of *VARD* scholarship, with their analytical aim based on the stylistic information that can be more effectively gathered with the assistance of the tool rather than in gauging the success of the normalisation process itself.

Normalising Middle English for stylistic examination is can be problematic, and is not without controversy. For one, there is a distinct degree of editorial conservatisim embedded in medieval textual studies which, as Markus (2000: 181-4) observes, tends to reject change in the form of substantive editorial intervention to the original text. Additionally, the Middle English in *PP* and its contemporary reference corpus contain

[35] *VARD* has also been used in corpus linguistic examinations of non-historical language characterised by spelling irregularity including including EFL learner corpora (Rayson & Baron 2011) and corpora containing electronically-mediated discourse (Tagg et al. 2012, Smith et al. 2014)

higher levels of spelling variation than the Shakespearian language normalised in Demmen (2012). While orthographically inconsistent, Shakespeare's late 16th century language was much more directly influenced by the influences of early-stage spelling standardisation than the 14th century English in *PP*, thus making it reasonable to expect a much higher level of spelling variation in Langland than anything written by Shakespeare.[36] This variant differentiation between late ME and EME is visible in Baron et al.'s (2009b) chart demonstrating the extent of non-regularised spellings across 6 historical language corpora, as reproduced below in Figure 2.10.

Figure 2.10: The percentage of variant types in historical Englishes chronologically by decade, as evidenced in *ARCHER*, *EEBO*, *Innsbruck*, *Lampeter*, *EMEMT*, and a corpus of Shakespeare's works (reproduced from Archer et al. 2015: 9, citing Baron et al. 2009b: 51)



Normalisation is moreover problematic from a literary linguistic perspective, for the changes to the text caused by lexical regularisation have the potential to minimise the

[36] For the differing amounts of orthographic variation present in Middle and Early Modern English, see Scragg (1974: 52-87).

significance (or entirely level) certain features of stylistic significance to the poem, particularly of a dialectal nature.

This investigation addresses these potential concerns raised by medieval studies and LL by orthographically annotating my corpora using XML tags. One of the options available in *VARD*, XML tags constitute a removable annotation structure that preserves the original spellings of the source-texts in the corpora while still achieving computationally-searchable text. This approach enables normalisation as additional method of analytical study that does not replace the original text (or the scholarship based on examination of said text). An example of this format is illustrated below with the *Gawan*/*Gawayn* variants highlighted in Figure 2.9, demonstrating how the spelling of the titular character's name *Gawan* is normalised to *Gawain* through this annotation scheme:

<normalised orig="Gawan">Gawain</normalised>

As noted by Baron and Rayson (2008: 11-12), the contents of these XML tags are designed to be overlooked by programs such as *WordSmith* when assembling word lists and determining frequency-based language patterns, and can be stripped away from the corpus as needed for further qualitative analysis. This is an important consideration for my analysis of style, which is interested in these unique linguistic features and distinctive spellings once corpus analysis sufficiently narrows the search parameters enough to facilitate stylistic analysis (via generating KWs for *PPCorp*). This dual capability of normalising orthographic representation without preventing lexical preservation is not possible using the plain-text format of annotation also offered through *VARD.*

My decision to normalise my two Middle English corpora is based on the belief that orthographic annotation improves the results of historical corpus stylistic analysis. Taking this approach acknowledges that complete normalisation of the corpus data is unlikely to occur (Markus 2000: 193, Baron & Rayson 2008), and also that using *VARD* will not solve all methodological difficulties associated with corpus analysis of Middle English text. Instead, the present study fundamentally aims to achieve more accurate frequency information in my corpus data than would otherwise be generated by using the original text for corpus analysis. Studies such as Smith et al. (2014: 150-1) suggest that the practical impact of regularisation on KW data may be minimal in terms of the differences in content and ordering achieved in post-annotated KW lists, but this conclusion is specifically addressed to studies working with variation in modern digitally-born language rather than historical data.

The current investigation primarily aims to use *VARD* as a method of cleaning the orthographic data in my target and reference corpora for further keyword and collocational analysis, but it is also used to conduct an initial exploration of the effectiveness the tool specifically in tackling Middle English spelling variation in poetry. This latter aim of ascertaining *VARD*'s success in reducing ME variants in its own right does not aspire to produce the level of statistically-based comprehensiveness achieved in studies like Rayson et.al (2007), Baron et.al (2009a), Baron et.al (2009b), or Baron et.al (2011), but rather sets out to provide a mixed qualitatively- and quantitatively-based description illustrating general trends in the annotated results and some of the identifiable factors contributing to the normalisation evidenced in my post-*VARD*ed late medieval corpora (see 3.3).

3. METHODOLOGY

My corpus approach to *PP* represents a pragmatic balance between the standards of corpus linguistic analysis, and the practical realities of studying a historical language. This chapter outlines how the considerations and challenges linked to my ME poetic source-texts (as set out in Chapter 2) are addressed in my chosen methodological approach. The decisions made in generating, categorising, and analysing the lexical data in *PP* take into account the poem's manuscript history, the historical nature of the language appearing in my corpora, and the precedents set by other CS investigations looking at non-Modern English literary texts.

The current methodological approach follows the example set by Rayson's (2008: 520) method of data-driven corpus analysis by 'first focussing on whole texts and then suggesting specific features to study in further detail'. Constituting one of several different approaches to corpus analysis, this 'Type III' methodology means that decisions regarding which linguistic features to study further (in this case, the KWs generated for *PP* addressing themes of consumption and the law) 'are made based on the basis of information extracted from the data itself', as set out in the process listed below (Rayson 2008: 521):

1. Build: Corpus design and compilation
2. Annotate: Manual or automatic analysis of the corpus
3. Retrieve: Quantitative and qualitative analyses of the corpus
4. Question: Devise a research question or model (iteration back to Step 3)
5. Interpret: Manual interpretation of the results or confirmation of the accuracy of the model

In this approach, initial corpus results are generated for *PP* (its word lists, keyword lists, and the frequency information contained in each) in step 3. The decision on what approach to take to said corpus data – and the dataset(s) to analyse further – directly ties back to the content produced in step 3, with the study's research focus devised

around the generated lexical results for *PP* after initial classification revealing the type of data to which I have access (in this case, through semantic-based KW categorisation). Once the linguistic features prominently evidenced in the categorisation have been identified (e.g. legal and consumption-based KWs), the process returns to step 3 for further quantitative and qualitative analysis of these narrowed datasets, which leads to the interpretation stage set out in stage 5. The interpretation stage within the current study includes both engaging with LL theoretical frameworks, as well as linking the linguistic analysis back to findings within the poem's literary criticism. [37]

The following chapter sets out the key decisions involved in taking this pattern-driven (e.g. contextually-informed but ultimately data-driven) approach to the lexis in *PP*. 3.1 introduces the *PP* corpus (hereafter, *PPCorp*) and its late medieval English poetic literary reference corpus (hereafter, *MEPoetLitCorp*), as well as the selection of each corpus' source-texts. 3.2 sets out the pre-analysis orthographic annotation performed on *PPCorp* and *MEPoetLitCorp* using the corpus tool *Variant Detector* (*VARD*), while 3.3 examines the methodological effectiveness of this regularisation using data drawn from both *PPCorp* and *MEPoetLitCorp*. 3.4 completes the chapter by outlining the processes by which the poem's keywords and collocational patterns are generated using *WordSmith Tools,* and signposts the statistical measures used and cut-off points established for determining significance within the generated corpus data.

[37] Here Rayson's (2008) methodological framework sets out the way in which Kopaczyk & Tyrkko's (2018) pattern-driven approach is being used in the current investigation as a data-driven approach characterised by the absence of expected theoretical neutrality.

3.1 Building the corpora

3.1.1 The corpora

This study makes use of two self-built specialised literary corpora. These corpora consist of:

(i) a target corpus containing the text for *Piers Plowman* (*PPCorp*)
(ii) a reference corpus of late Middle English poetic literature (*MEPoetLitCorp*)

PPCorp is comprised of a single textual version of *PP*, while *MEPoetLitCorp* contains a sampling of literature as represented by 28 late medieval poetic texts. I use the word 'sampling' here intentionally to acknowledge that historical language studies using corpora have difficulty achieving the levels of balance or representativeness traditionally expected in corpus linguistics (e.g. McEnery et al. 2006: 15-8), with the focus of creating *MEPoetLitCorp* instead based on the collection of a selection of digitally available literature against which comparison to *PP* is both linguistically and practically viable.

The majority of texts included in the reference corpus are thought to have been composed sometime between the mid 14th and early 15th centuries as poetic works contemporary to *PP*. My decision to extend the historical scope of the investigation beyond 14th century texts allows me to include a number of poems belonging to the *Piers Plowman* tradition in my reference corpus (Barr 1993). Their integration provides an additional measure of comparison against which to evaluate the similarity of language in *PP* to other related literary works, and provides a comprehensive view of the literary context surrounding the poem's style unconstrained by the artificial boundaries imposed by century markers. Given my primary interest in 14th century poetry, texts composed by other 15th century poets not associated with the *Piers*

Plowman tradition, including works by John Lydgate and Thomas Hoccleve, are not included in this investigation.

MEPoetLitCorp is divided into two smaller sub-corpora according to poetic form. While difficult to categorise historical texts, particularly those with complicated transmission and manuscript histories (Taavitsainen et al. 2006: 86), poetic form operates as one of the most distinctive features of ME literature through the division of texts into alliterative or metrical poetic traditions (2.2.3). Dividing the metrical verse and alliterative revival poetry into separate sub-sections of the broader reference corpus enables me to better observe the methods by which the language in *PP* engages with these two separate literary traditions, while still enacting a full comparison of the poem to the broader late Middle English poetic landscape (as represented by the surviving manuscripts from the time period). This approach is particularly useful in ascertaining whether features identified in *PP* depart from typical language use as register or style markers relative to the language norms present in each poetic tradition.

These sub-sections of *MEPoetLitCorp* containing metrical and alliterative poetry (hereafter *MVCorp* and *ARCorp*, respectively) are constructed as separately searchable components within the reference corpus, as shown in Table 3.1. The difference in size between the two sub-corpora in *MEPoetLitCorp* evidenced below is not representative of the number of poems contained in each part of the corpus, but rather of the comparative lengths of the poems present in each sub-corpus (for more discussion, see 3.1.2).

Table 3.1: Overview of the study's target and reference corpora

	Scope of coverage	**Abbreviation**	**Size by total word count (tokens)**
Target corpus	William Langland's *Piers Plowman*	*PPCorp*	70,541
Reference corpus	28 late ME poetic literary texts	*MEPoetLitCorp*	890,778
	Sub-corpus #1: Late ME alliterative revival poetry	*ARCorp*	374,807
	Sub-corpus #2: Late ME metrical poetry	*MVCorp*	515,971

The following section sets out the processes by which *PPCorp* and *MEPoetLitCorp* were constructed, including the selection of source-texts as informed by the poem's manuscript tradition and wider literary context, as well as the sources for the digital copies of said source-texts.

3.1.2 Corpus compilation

3.1.2.1 PPCorp

In building the target corpus I chose to use a digitised version of the C-text, which situates my research in one of the main manuscript traditions associated with the poem. As discussed in 2.1.2, the C-text constitutes one of the main authorial versions of the poem. My selection of the C-text in preference to the A- or B-texts, despite criticism of the version as being editorially incomplete (e.g. Pearsall 1994: 10), bases my stylistic analysis within the most chronologically recent surviving editorial adaptation of the poem.

The textual history of the C-text presents some interesting considerations for its use in stylistic research. According to Schmidt (1995a: xi), the C-text represents a

collection of 30 separate manuscripts: *X, Y, I, P^2, D^2, B, O, L, U, D, T, H^2, Ch, N^2, P, E, A, V, R, M, Q, W, F, Ca, S, Z, G, K, N,* and K^2 (Russell & Kane 1997: 1-18 also include *H* and *J*). These manuscripts share varying degrees of hereditary similiarity based on which manuscripts were copied from one another, and the text tradtions established as a result of these practices. The pervading critical view of the C-text and its manuscripts as less complete and thematically cohesive has resulted in heavy emphasis in textual critique and general literary scholarship being placed on the B-text in preference to C, a trend in *PP* scholarship until fairly recently with the rise of a new C-archetype.[38] In this context, an 'archetype' represents the manuscriptal version of the poem that most closely approximates the original, or authoritative, version of the text within its given manuscript tradition (A, B, C). The presence of a C-archetype is particularly notable due to the fact that the B-text, while deemed complete in terms of content-based cohesiveness and revisional completedness, does not exist as single archetypal manuscript but as a collated version of the text drawn from seventeen different existing manuscripts conforming to traditions defined by several 'sub-archetypes' and their related traditions (Kane & Donaldson 1975: 21). Using a collated version of the B-text such as the modern edition by Schmidt (1995c) – frequently adopted as the version of choice in modern university classrooms – would offer a 'reconstructed' version of the poem that did not actually exist in the 14th century, and thus a text to which no contemporary medieval reader would have had access. In this sense, using an archetypal C-text manuscript not only orients stylistic analysis around one of the oldest and most complete versions of the poem, but also bases analysis on a source-text that a late medieval reader could well have read as an original historical document.

[38] The relativity of this recentness is based on its later discovery in 1935, long after Skeat's (1888) seminal and heavily influential work on the editions of *PP* at the end of the nineteenth century.

The present study uses this archetypal version of the C-text, the *X* manuscript, as the source-text for *PPCorp*. Based in San Marino, Huntington Library MS 143 was first discovered by R. W. Chambers et al. (1935), wherein discussion of its potential as a C-archetype first transpires. Since the early twentieth century onwards, this copy has replaced Skeat's preference for *P* (Huntington MS 137) as the manuscript of choice for modern editions of the C-text including Salter and Pearsall (1969), Pearsall (1994), Russell and Kane (1997), and even in Schmidt's (1995a [2011]) parallel version of the different manuscript traditions.[39] Russell and Kane (1997: 175), while noting a relatively high number of scribal errors in the manuscript, highlight the qualities ultimately contributing to the sense of authenticity ascribed to *X* as aligning with the features expected to be present in the ideal *PP* copy-text:

> …to go by handwriting, it [X] appears to be among the oldest copies of C known to have survived. In a second feature, the approximation of its language to that of the author, it is also appropriate, for its West Midland dialect encompasses the poet's birthplace in Northern Shropshire. (clarification added)

A digital copy of *X* was accessed through the *Piers Plowman Electronic Archive* (hereafter, *PPEA*) (Bart et al. 2018). As briefly signposted in 2.1.5, the *PPEA* operates as a collaborative venture amongst *PP* scholars who volunteer their time and expertise towards producing edited electronic versions of as many textual versions of the poem as possible; this makes it possible to have access to multiple edited versions of the text simultaneously on an interactive digital platform. While the *X* manuscript was not yet available to the public in the early stages of this investigation (October 2015) nor yet at the time of submission (September 2018), the editors of the *PPEA* kindly provided an unedited, plain-text copy of the manuscript for use in this research.

[39] It is generally held in *PP* textual scholarship that Skeat did not have knowledge of the *X* manuscript, thus basing his conceptualisation of the C-text tradition around the next-best available version *P* (see Schmidt 1995b: 17).

It is worth briefly observing here that my selection of the *X* manuscript represents one of many stylistic analyses that can be performed on *PP*. The range of textual forms in which the poem appears has the potential to alter the results of a stylistic analysis on a case-by-case basis; a stylistic evaluation of the B-text will differ significantly from analysis of the C-text, with markers of style even potentially distinguishing different manuscript versions of the same textual tradition from one another (*X* from *P*, for instance). This comes before any additional considerations as to the number of ways in which the poem can be read and interpreted by different readers of the same textual version (Stockwell 2002: 3-5). What we can aim for in this type of investigation – and what I hope to achieve in using *X* – is an exploratory study of the style of the poem as drawn from the most authentic and editorally complete version of the C-text, an analysis informed by the empirical data generated through my chosen corpus methods.

3.1.2.2 MEPoetLitCorp

In contrast to the manuscript version chosen for *PPCorp*, my reference corpus is comprised of modern editions of these 28 medieval poetic texts. My use of *modern editions* in this capacity refers not to translations of the poems from ME to ModE, but rather ME versions that have been extensively edited by modern scholars to correct perceived errors and fill gaps in the manuscripts to facilitate improved engagement with historical literature by modern (and, at times, non-specialist) audiences. Using modern published editions as the source-texts for historical corpora follows previous approaches to building historical corpora as seen in, for example, Taavitsainen et al.'s

(2006: 80-1) *Middle English Medical Texts* (*MEMT*) corpus which uses a mixture of modern editions and manuscripts. It is also discussed in a literary stylistic context by Norgaard et al. (2010: 27-8), who also advocate using modern edited versions when studying historical style. Such decisions are generally based on reasons of practicality, for many of the ME texts being used for corpus analysis – be they literary or medical in nature – are only digitally available in machine-readable formats as modern editions; these modern editions can also evidence less severe spelling variation, depending on the level of editorial intervention applied to the text. This approach to corpus compilation necessarily prioritises collecting a large enough corpus of ME poetry from what material currently exists over arguments grounded in textual authenticity.

These source-texts for *MEPoetLitCorp* were acquired from several digital sources. Source-texts were drawn from general databases such as *Project Gutenberg* and the *Oxford Text Archive* (*OTA*), as well as more medieval-specific sites such as the *Corpus of Middle English Prose and Verse* (*CMEPV*) and University of Rochester's *Middle English Text Series* (*TEAMS*). *Project Gutenberg* contains versions of some of the more well-known poems (notably, those by Chaucer and the *Pearl* poet) but does not feature many of the lesser-known poems such as *Saint Erkenwald*, *The Awntyrs of Arthure,* and *The Pistel of Swete Susan*, which could only be found in the more medieval-specific databases. Where possible, I have chosen not to use source-texts from the *Internet Archive*; the formatting used by the archive appears to be of notably poorer quality than the versions offered by the other electronic resources. More information about the sites from which each source-text in *ARCorp* and *MVCorp* was drawn from can be found in Tables 3.2 and 3.3, respectively (see Appendix A).

Selection of which poems to include in each sub-corpus in *MEPoetLitCorp* was first determined by their inclusion in the study's pre-specified range of composition

dates between 1350 and 1500. While not always possible to establish concrete dates (or, in reality, a specific year) in which a given text was written, these poetic origins have in most cases been localised to a span of years or decade in which composition was most likely to have occurred through analysis of poetic technique, types of regional vocabulary use, and other stylometric parameters. Identifying approximate composition dates is particularly difficult for poetic texts not associated with a historically-identifiable authorial figure; the poems written by the *Pearl* poet (*Sir Gawain*, *Pearl*, *Patience*, and *Cleanness*), for instance, can only be generally attributed as late 14th century works dating somewhere between 1360-1395. A few texts such as *Joseph of Arimathie* and *Gologras and Gawain* are considered to have been composed on the respective outer fringes of the 1350-1500 time period, but have been included nonetheless based on their presence within the general late medieval literary landscape as integral parts of the alliterative revival tradition (Turville-Petre 1977: 146).

The second parameter for source-text selection involved the more practical assessment of its digital availability. As discussed in 3.1.1, the construction of *MEPoetLitCorp* was aimed not at including all surviving late medieval English poetic texts fitting within the specified compositional dates but was rather meant to provide an approximate representative sampling of material for exploratory corpus analysis (McEnery & Hardie 2012: 8-9). To this end, the omission of texts composed within the targeted temporal window but not available in machine-readable electronic formats at the time of corpus construction (October 2015) such as *Alexander of Macdeon*, *Alexander and Dindimus,* and *The Crowned King* is not seen as a flaw in the corpus makeup but rather an indication of present limitations on data accessibility when examining medieval literary material.

Despite the lack of availability of certain source-texts, several late ME poems appear in multiple digital depositories simultaneously. These texts are characteristically attributed to the most well-known authors (anything by Chaucer, for instance) or the most recognisable of the alliterative revival texts, including the alliterative *Morte Arthure* and *The Siege of Jerusalem.* The high levels of scholarly interest paid to these particular poems may well account for their availability in several different editions, and from multiple electronic archives; there is also strong correlative evidence between digital availability and the number of surviving manuscripts per individual work. In these cases selection of the chosen digital copy for the reference corpus' source-texts was based on a) the quality of digitisation, and b) the quality of the transcription evidenced compared to their original versions. In instances of seemingly equal quality, efforts were made to use sources compiled by the same modern editor when handling multiple texts attributed to the same author, even if not the most mainstream or 'authoritative' versions as evidenced within the relevant literary criticism.[40]

The alliterative sub-corpus in *MEPoetLitCorp* (*ARCorp*) is comprised of eighteen ME alliterative poems, as set out below in Table 3.4. Poems are listed alphabetically by title, with the column titled 'tokens' indicating the complete word count of each text after cleaning the text files in preparation for corpus use (see end of present section). The titular abbreviations listed here are those assigned to each text, which are used for the remainder of this investigation.

[40] For an example of this editorial uniformity, see the use of Skeat (1923a, b, c, and d) for Chaucer's texts in Table 3.3 in Appendix A.

Table 3.4: List of the alliterative poems in *ARCorp* by title, abbreviated reference, and text length by total word count.

Text	**Abbreviation**	**Size (tokens)**
The alliterative Morte Arthure	*aMA*	37,457
The Awntyrs of Arthure	*AoA*	5,338
Cleanness	*Cl*	16,412
The Destruction of Troy	*DT*	121,665
Gologras and Gawain	*GG*	10,482
Joseph of Arimathie	*JA*	6,384
Mum and the Sothsegger	*MS*	16,232
The Parlement of the Thre Ages	*PTA*	6,205
Patience	*Pa*	4,986
Pearl	*Pe*	8,421
Pierce the Ploughman's Crede	*PPC*	7,648
The Pistel of Swete Susan	*PSS*	2,515
Richard the Redeless	*RR*	7,616
Saint Erkenwald	*SE*	3,505
The Siege of Jerusalem	*SJ*	11,548
Sir Gawain and the Green Knight	*SG*	20,996
The Wars of Alexander	*WA*	82,651
Wynnere and Wastoure	*WW*	4,746
		TOTAL: 374,807

While not possible for a medieval corpus to provide balance in a modern corpus linguistic understanding of the term (i.e., Biber 1993), the texts in *ARCorp* provide a chronologically comprehensive representation of the alliterative revival. Coverage begins with earlier poems such as *Wynnere and Wastoure* (*WW*) and *Joseph of Arimathie* (*JA*) starting in the mid-14th century, and ends with *The Awntyrs of Arthure* (*AoA*)*, Gologras and Gawain* (*GG*)*,* and the alliterative *Morte Arthure* (*aMA*) in the early to mid-15th century. The majority of poems in the corpus, including the *Pearl* poems and *Saint Erkenwald* (*SE*), are generally clustered somewhere within the late 14th century; this places the linguistic orientation of *ARCorp* within an approximate 50 year time frame between 1350 and 1400. The corpus also includes alliterative

representatives of the *Piers Plowman* tradition in *Mum and the Sothsegger* (*MS*), *Pierce the Ploughman's Crede* (*PPC*), and *Richard the Redeless* (*RR*).

By comparison, *MVCorp* consists of ten texts stemming from a more general poetic tradition of metrical verse in late medieval English. The works of two primary authors associated with 14th century English poetry, Chaucer and Gower, feature prominently in the metrical poetic landscape and are represented accordingly within *MVCorp*.

Table 3.5: List of the metrical poems in *MVCorp* by title, abbreviation, and text length by total word count

Text	**Abbreviation**	**Size (tokens)**
Anelida and Arcite	*AaA*	2,790
The Book of the Duchess	*BD*	8,628
The Canterbury Tales	*CT*	133,985
Confessio Amantis	*CA*	207,382
The House of Fame	*HF*	13,223
The Legend of Good Women	*LGW*	21,659
The Parlement of Foules	*PF*	5,524
The Plowman's Tale	*PT*	8,827
The Romance of the Rose	*RotR*	48,401
Troilus and Criseyde	*TC*	65,552
		TOTAL: 515,971

As evidenced above in Table 3.5, Chaucer is heavily represented in the number of texts present within *MVCorp* (eight out of ten texts), reflecting his status as the most prolific writer of the late 14th century. Almost all of his works are included in the corpus, with the exception of his prose *Treatise of the Astrolabe* and some of his shorter poems. Gower, while only authoring a single text in the corpus, is a significant presence in *MVCorp* on account of the sheer size of his poem *Confessio Amantis* (*CA*) which, at

over 200,000 words, constitutes nearly 40% of the corpus by itself. As seen in *ARCorp*, the *Piers Plowman* tradition is also represented with the inclusion of *The Plowman's Tale* (*PT*).[41]

Together, *ARCorp* and *MVCorp* constitute a collective linguistic backdrop that, for the purposes of the present stylistic investigation, serves as a relative norm of late medieval poetic language use against which to compare the language occurring in *PP*. The presence of 28 source-texts in *MEPoetLitCorp* aligns my investigation with other corpus stylistic reference corpora, including Mahlberg's (2013) corpus of 29 19th century novels and Demmen's (2012) reference corpus of 43 non-Shakespearian early modern plays. My reference corpus departs from these examples in that its tokens, or total word counts, are considerably smaller; I would argue that this is indicative of the text-type being studied, rather than signifying insufficient coverage within the corpus itself.

Prior to use in the reference corpus, the quality of these digital source-texts was appraised for consistency with the original versions of the poems. This was achieved by comparing the source-texts to their corresponding print copies to ensure the textual accuracy of the transcription, using the first 200 lines of each poem as a baseline for this quality-assurance testing. This step was aimed at minimising the level of transcription-based error that could potentially influence my corpus results.

After source-text selection, these texts were subject to pre-annotation file scrubbing to remove extralinguistic material and content. Corpus programs like *WordSmith* do not have the capacity to distinguish the scholarly material present in these modern editions from the actual literary text, a limitation in computational

[41] The remaining texts of the Piers Plowman tradition (*The Prayer and Complaynte of the Ploweman unto Christe* and the texts surrounding *Jack Upland* and its reception) are written in prose, and for this reason are not included in the reference corpus.

awareness which necessitates the manual removal of this extraneous information before any corpus data can be generated. To this end all editorial content – including orthographic notations, modern-day glosses of difficult words and phrases, indexes, glossaries, and any other type of scholarly commentary – was manually removed from each source-text in *PPCorp, ARCorp,* and *MVCorp*. Where relevant, the texts within these corpora were also divested of structural information such as subheadings, line numbers, and folio demarcations.

These stripped-down versions of the poems were then saved as plain-text (.txt) files as the format accepted by both *VARD2* and *WordSmith*. Texts with distinct sections like Chaucer's *Canterbury Tales* were saved individually (e.g., the *Knight's Tale* and the *Miller's Tale* constituting individual searchable elements within *MVCorp* as separate from their respective prologues), bringing the total number of text files for the 28 poems in *MEPoetLitCorp* to 76.

3.2 Annotating the corpora

Prior to corpus analysis the source-texts for *PPCorp* and *MEPoetLitCorp* were also annotated for orthographic variation. This approach follows the example set by studies like Culpeper and Kyto (2010) and Demmen (2012) in using *VARD* as a pre-processor for corpus (stylistic) analysis in historical text. While these investigations (Demmen 2012 in particular) focus on the EME used by Shakespeare and his literary contemporaries, it is not difficult to foresee how taking a similar approach would benefit analysis of the late ME literature appearing approximately two centuries earlier. The fact that the tool was initially designed to assist studies in EME (Archer

et.al 2003) does not hinder investigations of ME but rather signifies that a certain degree of caution be extended in the use of this tool.

My use of *VARD* constitutes a two-step process, with the first stage aimed at manually training the tool to more accurately engage with the ME poetry being annotated. This training involves working with a selection of text drawn from *PPCorp* and *MEPoetLitCorp*, and occurs through the manual annotation interface built into the tool's training scheme (Baron & Rayson 2008: 10). The second step of orthographic normalisation involves the automated annotation of the entire target and reference corpus, a task *VARD* facilitates by utilising the information extracted from the manual training. I set out the parameters of this manual training in 3.2.1, then discuss the subsequent automated annotation of *PPCorp* and *MEPoetLitCorp* in 3.2.2.

3.2.1 Manual annotation

As outlined in 2.4.1, the fundamental aim of orthographic annotation in the present study is to reduce the amount of spelling variation evident in my Middle English corpora. This reduction is quantitatively trackable by progressively measuring the number of *types*, or total number of individual word forms (as opposed to tokens as the total word count in a corpus), contained in each corpus throughout the annotation process. These type-based parameters apply to both stages – manual and automated annotation – of this regularisation process.

To achieve this overarching aim, the initial manual training stage focused on a selection of language drawn from across *PPCorp* and *MEPoetLitCorp*. To determine the amount of text needed to train *VARD* on Middle English I follow the findings in Baron and Rayson (2009) suggesting that the use and overall effectiveness of *VARD* in

offering accurate replacements improves in proportion to the amount of linguistic input it receives to certain thresholds, with:

> the first 2-3,000 tokens of training yield[ing] a steep increase in performance… after this a not quite so steep increase in performance continues until 10-12,000 tokens. After this stage, each 1,000 tokens only increases the tool's performance by a small amount so extra training is not really of significant enough benefit (Baron & Rayson 2009: 21).

Developed from a study exploring Early Modern English as well as the variation present in children's language, these findings set out an approach to orthographic training that is cost effective in terms of time and effort relative to the measurable benefits further annotation affords.

In recognition that the tool's ability to learn plateaus after a certain amount of manual training (Baron & Rayson 2009: 9-22), I originally trained *VARD* using approximately 16,000 tokens. This decision situated my approach in relatively close proximity to the ballpark figure of 12,000 given by Baron and Rayson (2009: 21) as the peak rate of productive annotation for EME where the value of the tool's further training starts to progressively diminish. I followed Demmen (2012) in setting my cut-off point slightly higher than Baron and Rayson's (2009) standard, my reason for doing so based on the more historical nature of the language present in my corpora.

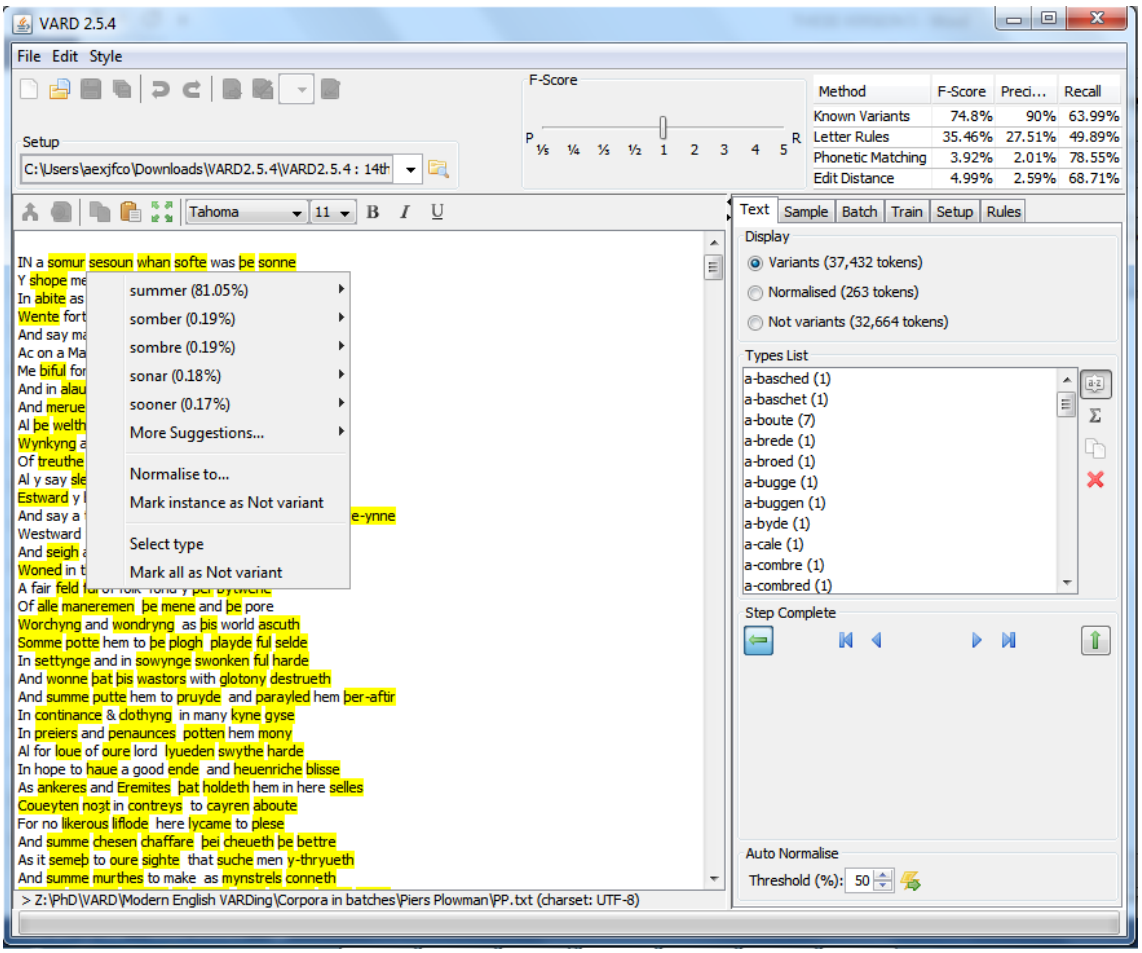
The annotation of these 16,000 tokens occurred across the poetic source-texts in *PPCorp* and *MEPoetLitCorp*. On the advice of Baron (in person, 2015), I divided the text of my corpora into 'batches' of 250 words within *VARD*, which provided small datasets of language on which to manually train the tool. These batches were selected at random by *VARD* for my attention to encourage general ME normalisation across the literary dataset, a move aimed at preventing any underlying bias from the analyst in inadvertently prioritising different types of language (e.g. alliterative vs. metrical) or

otherwise tailoring normalisation to the spelling conventions typifying a particular author or regional location during the selection process.[42]

Within each batch, I was able to normalise instances of spelling variation identified by *VARD* by clicking on each individual variant (highlighted in yellow as shown in Figure 3.6 below) and either selecting the most suitable replacement form offered or providing a replacement form not offered by the tool. My annotations within these individual batches familiarised the tool with the type of spelling variation typically seen in my late ME text, an awareness which gradually manifested in the better selection of 'replacement forms' being offered for these variants. These replacement forms originated from the Modern English word list contained in *VARD*; it is this departure of a word form from this word list that prompts its identification as a spelling variant (Baron 2008).

[42] For more discussion about the impacts of this randomised batch selection on the resulting normalisation, see 3.3.3.

Figure 3.6: Screenshot illustrating VARD's training interface wherein the manual training of *PPCorp* and *MEPoetLitCorp* occurred. This image uses the batched, but non-annotated version of *PPCorp* to signpost not only the amount of spelling variation present in these ME texts, but also to demonstrate the interface by which these variants were normalised in the programme via the first 35 lines of the poem.



To this end, the normalisation of the language in my corpora occurred through the modernisation of the ME word forms to their Modern English orthographic forms. An initial attempt was made to regularise the language in my corpora to the headword entries in the *Middle English Dictionary* (using a word list provided by said resource). However, this did not prove effective in lowering the number of variants in the corpus through reduced token counts, so this approach was subsequently abandoned in preference for modernisation.

After manually training my corpora on 16,000 words, however, I was obliged to manually annotate another 16,000 exclusively from *PPCorp*. This decision was based on the fact that only 2 out of the 64 batches were drawn from the text of *PP* (see Figure 3.13). Given the uniqueness of *PP*'s orthography compared to other contemporary texts in terms of its regional dialect alone (e.g. Samuels 1985, Duggan 1990), very little of the normalisation conducted based on the texts in *MEPoetLitCorp* (i.e., the other 62 batches) impacted the language in *PP* after a brief survey of the word lists generated for *PPCorp* following automated normalisation. In other words, many of the spelling variants in *PP* were too unusual compared to the variant word forms encountered in the reference corpus to enable *VARD* to transfer or otherwise apply the knowledge to the text in *PP*, even If the word had already been encountered in one of the batches for a text in *MEPoetLitCorp.*

This additional step meant that the manual training undertaken in this study was comprised of two 16,000 word training phases, with the second phase dedicated exclusively to the language contained in *PPCorp*. This resulted in *VARD* being trained using 32,000 total words of Middle English across 128 250-word batches, and produced a version of the tool that was not only trained in the types of language appearing in late ME poetic literature but specifically on the types of language found in *PP*. Having these two phases of manual annotation, in addition to the higher number of words in the training data, mirrors the approach to othrographic regularisation taken to the corpus *of Early Modern English Medical Texts* (*EMEMT*) set out in Lehto et al. (2010: 285), which does contain some late Middle English text starting from 1375 onwards.

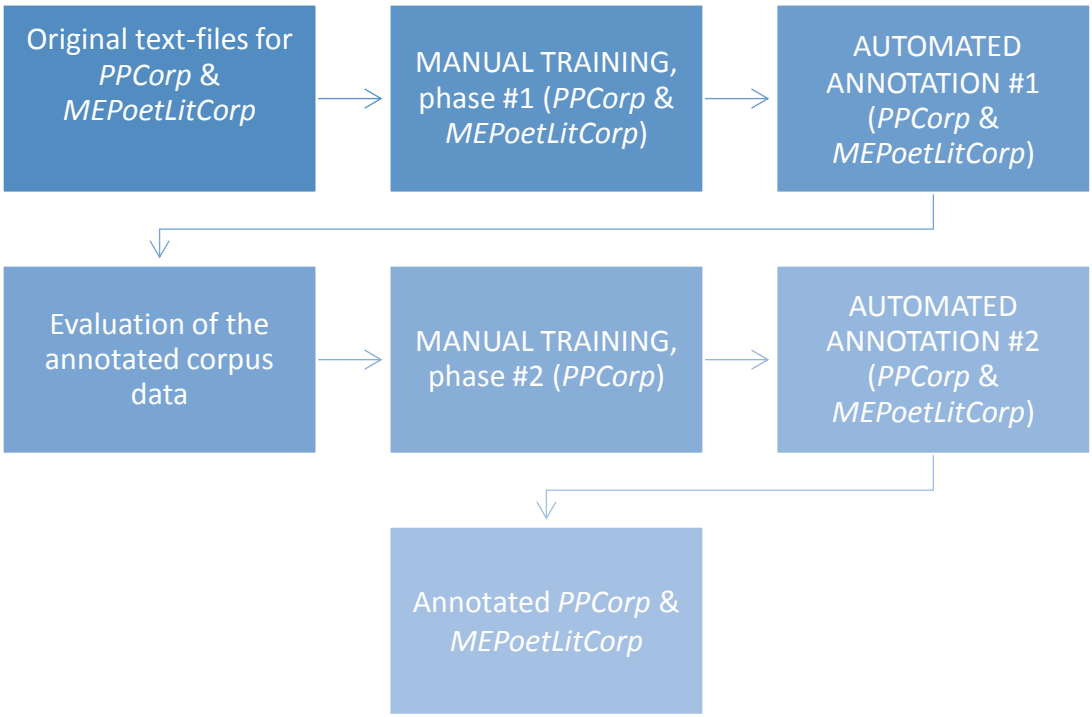
In conducting this manual annotation I followed the normalisation guidelines set out by Archer et al. (2015). Collated from existing corpus studies on orthographic

annotation of historical language, Archer et al. (2015:17-19) recommend that the annotation of historical textual data focus on specific groupings of features shown to either directly benefit from annotation including genitives and auxiliaries, or else treat with caution like compounds and archaic terminology (2015: 12-16). These features stand in contrast to terms that cannot be trained for automated annotation with tools like *VARD*, and rather rely on contextual clues for correct regularisation into its modern English form (e.g. ME pronouns).

3.2.2 Automated annotation

The second stage of this normalisation process marked a shift from manual to automated annotation. This regularisation performed by *VARD*'s automated interface was based on the knowledge accumulated from the two phases of manual training, and was applied to *PPCorp* and *MEPoetLitCorp* following each round of manual training as illustrated in Figure 3.7 below.

Figure 3.7: Diagram visualising the orthographic annotation process conducted on *PPCorp* and *MEPoetLitCorp* prior to corpus analysis



Although *VARD* was not trained on the reference corpus during the second phase of manual training, running *MEPoetLitCorp* through the tool's automated tagger a second time led to greater overall reduction in the number of types in the corpus (see 7.4.2.1). This potentially suggests that the continued training not only improved the tool's ability to handle variant spelling in *PP* but in ME poetic literature more generally, though more quality-based information about variant recognition and substitution is needed before making any definitive claims.

To produce this type of corpus-wide automated annotation, I uploaded the batched text-file versions of *PPCorp* and *MEPoetLitCorp* used in the manual annotation into the modernisation-oriented trained version of *VARD*.[43] The resulting annotated output was orthographically tagged in such a way that corpus tools such as *WordSmith*

[43] This constituted an important step, for my downloaded copy of the tool contained two separate versions, one trained to normalise to Middle English (which was ultimately not used in this investigation), and the one regularising variants into their Modern English forms.

would only process the normalised forms rather than the underlying variants. Words that remained unchanged during the annotation process (e.g. not identified as variants or homographs with an existing Modern English word) were still able to be extracted from the corpus by *WordSmith* normally without any residual problems resulting from the corpora's overarching orthographic tagging.

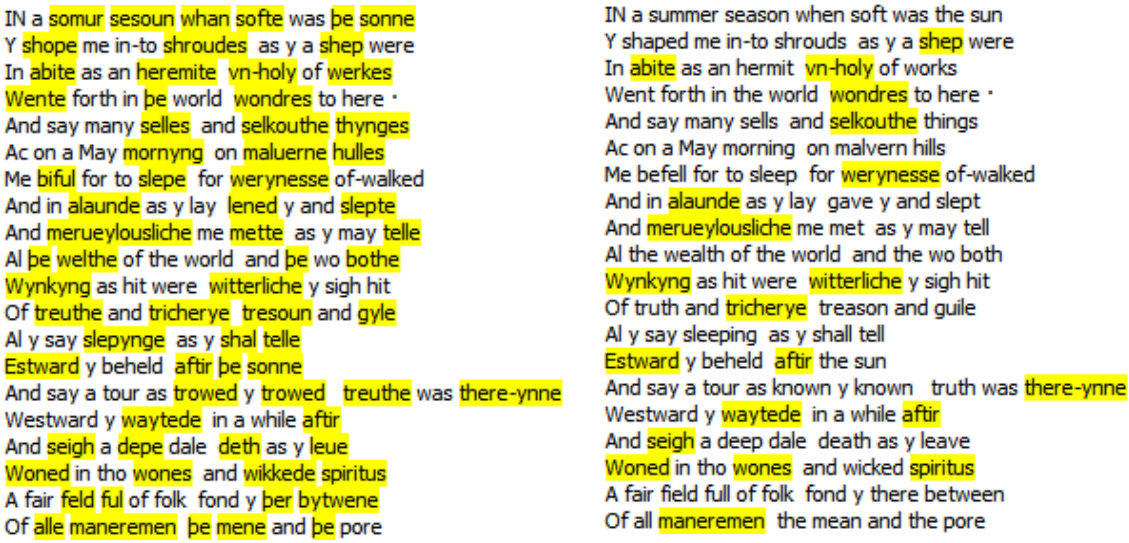
My use of *VARD*'s automated annotation tool – and the normalisation resulting from said annotation – is directly based on the confidence threshold selected for this investigation. As mentioned briefly in 2.4.1, confidence thresholds or 'replacement thresholds' (Baron & Rayson 2009: 15) refer to the percentage of confidence *VARD* attributes to the connection between an identified variant and its offered replacement form, with automated annotation only occurring if the tool's confidence percentage is higher than the threshold percentage established the by researcher. As outlined by Baron and Rayson (2009: 11) and Lehto (2010: 284), higher replacement thresholds result in less overall annotation throughout the corpus but the resulting annotation is generally more accurate, while setting lower thresholds loses accuracy but facilitates greater amounts of overall normalisation. Given the lack of time to experiment with the tool's settings to determine the best threshold for my Middle English text as suggested by Baron and Rayson (2009), I chose to replicate the 50% confidence threshold used by Demmen (2012: 151) which also serves as the default setting in *VARD.*

3.3 Impact of regularisation

Using *VARD* ultimately provided me with two versions of each literary corpus. The orthographically tagged text for *PPCorp* and *MEPoetLitCorp* (hereafter referred to as

the ‘regularised’ or ‘normalised’ versions) were produced through the cumulative annotation conducted in both the manual and automated tagging stages; these versions represent outputs of late medieval text normalised to a greater degree than evidenced in the original corpora, or the ‘non-normalised’ versions. Comparative qualitative assessment of the improvement in the normalised *PPCorp* over the original text is exemplified via the reduction of identified spelling variants after regularisation, as illustrated below in Figure 3.8. The batch chosen here (the first 250 words of *PP*) was not one of the manually trained text excerpts but rather achieved the resultant spelling regularity via automated normalisation.

Figure 3.8: Comparative visualisation of the spelling regularity contained in *VARD*ed (right) versus non-*VARD*ed (left) ME text in the first 20 lines of the normalised version of *PPCorp,* with variants highlighted in yellow.



As anticipated in 2.4, the combined approach of 32,000 words of manual training on – and the subsequent automated annotation of – my ME literary corpora produced some but not completely orthographically normalised text. This resulting output was characterised by the presence of modernised spellings alongside original

ME variants, a state of partial modernisation visible not only in Figure 3.8 but across both normalised datasets. This result does not suggest a failed method but rather follows other historical corpus studies wherein residual non-modernised word forms are seen not only as normative but also to some degree inevitable (e.g., Markus 2000: 193, Baron & Rayson 2008, Lehto 2010: 289).

To fully demonstrate the effectiveness of this pre-processing step in the present investigation, an abridged examination of the post-annotated datasets is set out below in 3.3.1. This reflexive mini-study identifies analytical benchmarks as a point of departure for critically assessing parameters such as success, accuracy, and saturation by comparing the regularised output of *PPCorp* and *MEPoetLitCorp* to the original text after both manual and automated training. While not a comprehensive examination, the aim of following this line of inquiry is to answer (v-viii) in RQ2 by observing some of the leading factors influencing my corpus-based annotation of ME text in order to better identify areas of improvement in future attempts at normalising medieval literary corpora, and anticipate how these will impact further CL analysis.

3.3.1 Benchmarks of effectiveness

One of the issues with discussing 'effectiveness' in corpus methods like orthographic normalisation is the inadvertent tendency to amalgamate success in using the method with the quality-based accuracy of the annotation itself. To this end, my current approach examines three measures of effectiveness: token reduction, variant replacement, and amounts of regularisation. Token reduction measures the total overall decrease in *types*, or the number of unique word forms in a corpus, to ascertain success in using annotation to minimise variation, regardless of annotational quality.

Variant replacement compares the normalised variants to their original forms on a case-by-case basis to offer commentary on the quality of the annotation taking place. Amounts of regularisation, by contrast, assesses the overall amount of normalisation (in this case, modernisation) present in a regularised text or corpus, which provides a parameter through which to explore the overall impact of annotation.

3.3.1.1 Type reduction

The decreasing numbers of types is first evidenced in the manual annotation stages. Table 3.9 below shows the gradual yet progressive loss of variant types demonstrated in readings taken in 2,000-word increments during the first stage of manual annotation (i.e., the first 16,000 words). This loss of types, ostensibly due to the lemmatising-esque effect produced in the merging of multiple ME variant forms under a single ModE form, is shown to have occurred within both *PPCorp* and *MEPoetLitCorp,* as well as the separate poetic traditions within the reference corpus (*ARCorp* and *MVCorp*). The listed type amounts were acquired by using *WordSmith* to produce word lists for *PPCorp, ARCorp, MVCorp*, and the combined *MEPoetLitCorp* after each 2,000-word block of manual annotation and counting the number of lexical entries listed for each (sub)corpus.

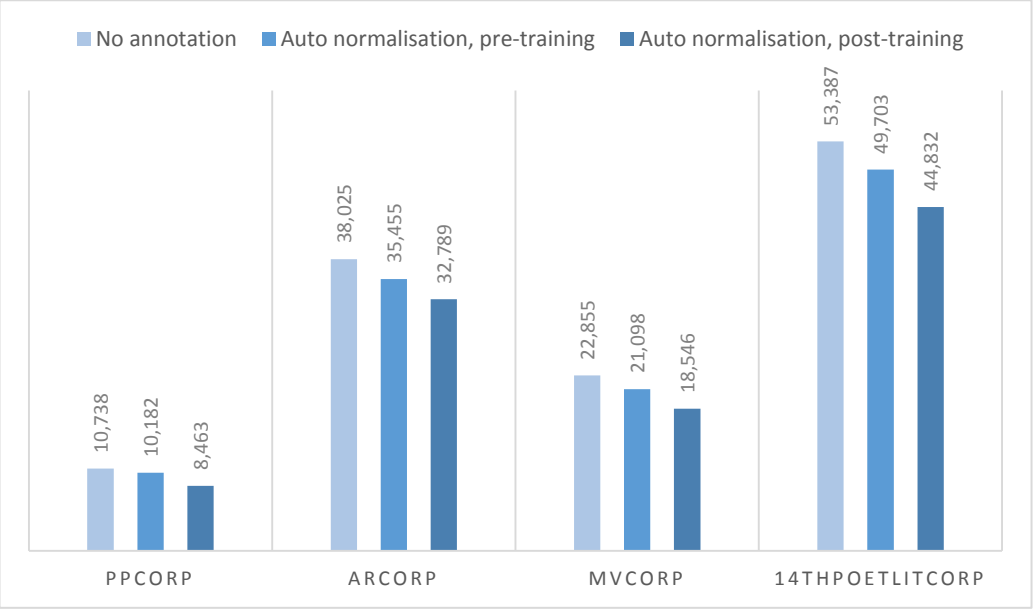
Table 3.9: Quantitative evaluation of annotation results as determined by progressively reduced numbers of types for each corpus; amounts measured along 2,000-word increments (every 8 batches of 250 words each).

	PP	ARCorp	MVCorp	both corpora
Original text	10,738	38,025	22,855	53,387
No manual VARDing	10,182	35,455	21,098	49,703
VARD, 2000	10,182	35,463	21,106	49,715
VARD, 4000	10,142	35,342	21,016	49,564
VARD, 6000	10,122	35,197	20,938	49,370
VARD, 8000	10,089	35,049	20,846	49,184
VARD, 10000	10,061	34,959	20,769	49,044
VARD, 12000	10,037	34,869	20,673	48,894
VARD, 14000	10,020	34,730	20,633	48,728
VARD, 16000	9,995	34,651	20,568	48,609

While duplicate records were not maintained for the second stage of manual training (16,000 words on *PP* exclusively), a similar trend was evidenced in the decrease of overall types generated for *PPCorp* by the end of this stage as well.

The quantifiable difference between pre-annotation and post-annotation types is also measurable in the corpora's automated annotation in *VARD*, as illustrated below in Figure 3.10. As with Table 3.9, data for the reference corpus has been divided into its alliterative and metrical sub-corpora (*ARCorp* and *MVCorp*, respectively) to show the type reduction evidenced within each of the contained poetic text-types. The 'no annotation' values in the graph (lightest blue) signify the total number of types in each corpus in the original text, while the 'pre-training' category (mid-blue) contains the number of types calculated as a result of running the corpora through *VARD* without any manual training. The final category 'post-training' contains the number of types generated for each corpus after a) the combined stages of manual training and b) the auto-normalisation of the corpora based on said training (dark blue).

Figure 3.10: Demonstrated effect of annotation on type reduction within my poetic corpora, as compared to the original and non-trained versions.



As evidenced above, the number of types are shown to have fallen for both the target and reference corpora as a direct result of orthographic annotation (see also Table 7.12). This progressive reduction of types illustrated in Figure 3.10 ultimately corroborates the stance taken by Baron & Rayson (2009) that a dual manual and automated integrated approach is more beneficial for the orthographic annotation of historical text over the use of each interface separately.

3.3.1.2 Variant replacement

While type reductions signal methodological success in fulfilling the aims of regularisation in the present study, this information does not provide any indication of the accuracy of these replacements. Accuracy represents a particularly problematic aspect of annotation, both manual and automated. Concerns with the accuracy of manual annotation tend to revolve around subjectivity (Sinclair 1992), while automatic

annotation is unlikely to be 100% error-free and will invariably produce results not likely to be seen in (or enacted though) manually-conducted annotation (Hunston 2002: 91).

Qualitative benchmarks of assessing annotation-based effectiveness are accessible by comparing the regularised forms to their original variants. Constraints in time and scope necessarily prevent a systematic appraisal of the resulting quality of normalisation through detailed sampling of the post-annotation batches to assess which replacements had been made and the accuracy of these replacements, as this would constitute its own study in its own right. However, a cursory examination of sampled batches of the annotated data suggested surprisingly high occurrences of error in the final corpora; taking the mean value of these samples and applying it broadly to the annotated corpora suggests that the overall project achieved a ballpark accuracy (i.e., precision) of approximately 42-47%. This figure did not significantly impede corpus analysis in the present investigation because every keyword, collocational pattern, and concordance line being analysed was compared back to the original text, but it does stand in marked contrast to the 92% precision score reached for EModE texts achieved by Baron and Rayson (2009: 14-18) when using a 50% confidence threshold.

From amongst the surveyed variant replacement pairings, measures of accuracy can be broken down into distinctions based on decisions made by the tool versus the human annotator. To reflect the differences in each type of process, decisions made by the annotator fall under the remit of ‘subjectivity’ while outcomes resulting from the tool are better defined as ‘limitations’. The effects of both this human subjectivity and technological limitation on regularisation are set out below in Table 3.11.

Table 3.11: Select examples of challenges to accuracy of *VARD*ing, as divided into tool-based limitations and decision-based subjectivity by the annotator

	Annotator	Tool
General setup of annotation in current project	Confidence threshold Amt. of manual annotation Letter replacement rules	
Manual annotation	Archaic/obsolete words Semantic hierarchy Polysemy/homonymy Erroneous annotations	False cognates Polsemy/Homonymy Abbreviations Variant identification
Automated annotation		Confidence threshold

The factors and language features listed in this table are not comprehensive but signpost consistent patterns observed in the post-annotated corpus data, with a more thorough exploration of textual examples for each unfortunately falling outside the remit of the present discussion.

3.3.1.3 Amounts of modernisation

By contrast to the quality-based accuracy revealed by the variant replacements, the amounts of annotation present in the regularised corpus materials reveal the broad impact of annotation on my ME corpora. This measure is necessarily estranged from precision in preference for observed saturation of modernised forms, which changes the scope of this measure from individual language features to broader patterns demonstrated in the presence or absence of normalisation across large sections of the

given corpora. This section thus focuses on dominant trends and some of the leading factors responsible for their presence in my corpus data.

An example of comparative saturation can be seen in Table 3.12, which illustrates the different levels of modernisation present in textual extracts taken from *Confessio Amantis* (CA), *The Miller's Tale* (*MilT*), *Joseph of Arimathie* (*JA*), *Sir Gawain and the Green Knight* (*SG*), *The Wars of Alexander* (*WA*), and of course *Piers Plowman* (*PP*). These titles have been selected from across both (sub)corpora to get a broad picture of annotation-based impact evidenced within the given corpus materials. Each passage, while randomly selected, was reviewed in the context of its originating source-text to ensure that the variant/replacement ratios present in these excepts was reflective or otherwise representative of the amounts observed in the work more generally.

Table 3.12: Randomly-selected extracts drawn from a selection of post-annotated texts in *PPCorp* and *MEPoetLitCorp*, with remaining variants highlighted in yellow.

Passage	Text
All other beasts that men find They serve unto here own kind, But to reason the Soul serves; Whereof the man his thonk deserves And get hem with his works good…	*CA*
This Carpenter had wedded new a wife Which that he loved more than hise life; Of eighteen year she was of age. Ialous he was, and held hir narwe in cage, For she was wild and young, and he was old And deemed himself ben like a cokewold.	*CT - MilT*
"Now I þonke my lord," said Ioseph "that lent me of his grace; me þinkeþ but þreo ni?t al this same þrowe." then Ioseph asks fontston & is I-folwed blyue; they cam after hem and his wife & with hem ful many.	*JA*
Then commanded the king the kny?t for to rise; And he full radly vpros, and ruchched him fair, Kneeled don bifor the king, and cachez that weppen; And he luflyly hit him laft, and lyfte up his hand, And gef him Goddez blessing, and gladli him bids That his hert and his hand schulde hardi be both	*SG*
They have tald me be-for this time that now I true find, How I should dwell in a land and look on a qwene." Then out of his bosom he brayd a blesand table Of Euour and of other þingis oddly formed, Of bres and of burnt gold and of bri?t silver…	*WA*
There preached a Pardoner as he a priest were And brouth forth a bull with bischopis selys Said that himself might assoylen hem all Of falseness and faystynges of vows ybrokene Ignorant men leued him well and liked his words	*PP*

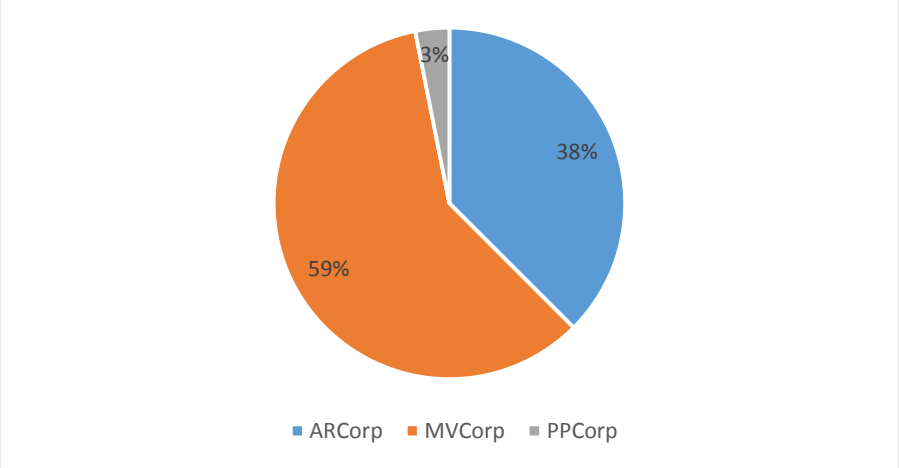
As demonstrated above in Table 3.12, the highest amounts of overall regularisation – signposted via the absence of high textual densities of remaining variants – appear in *CA* and *MilT*. These titles still contain periodic variants not

regularised during the annotation process but their text is generally characterised by possessing greater amounts of modernisation than historical word forms. *WA* and *PP* follow these leaders at distance, with their texts holistically comprised of approximately equal parts Modern and Middle English, depending on the given sections being viewed in each poem. By contrast, the passages for *JA* and *SG* are suggestive of the overwhelming presence of historical variants still present in the post-annotated text for these titles.

Several factors are potentially responsible for the variation demonstrated in the annotated output for these texts. One of the factors likely to have had an impact on this disparity was the dispersion of the manual annotation across the target and reference corpora in the first manual training stage (i.e., the first 16,000 words with which *VARD* was trained). Since batch selection was randomised it was not possible to ensure equal distribution in the amounts of annotation within the corpus materials, which resulted in disproportionately-balanced coverage of each (sub)corpus. This uneven contact in training is reflected in Figure 3.13, which shows the overall percentage of batches selected for manual training by corpus. Out of 64 total batches of training conducted in this stage, textual excerpts from *MVCorp* constituted 38 of these randomly-selected batches while only 24 of these batches originated from texts in *ARCorp*. These both dominated *PPCorp*, which was responsible for 2 batches, or 3% overall, of the first stage of annotation performed on these poetic corpora. This resulted in in more direct training on the metrical verse poetry (e.g., *CT* and *CA*) over the alliterative revival texts (*SG, WA, JA*), including *PP*.[44]

[44] This comparatively small focus on *PP* was also responsible for the decision to train *VARD* using an addition 16,000 words pulled exclusively from the poem (as set out in 3.2).

Figure 3.13: Corpus-based distribution of the batches randomly selected for the first 16,000-word manual training stage



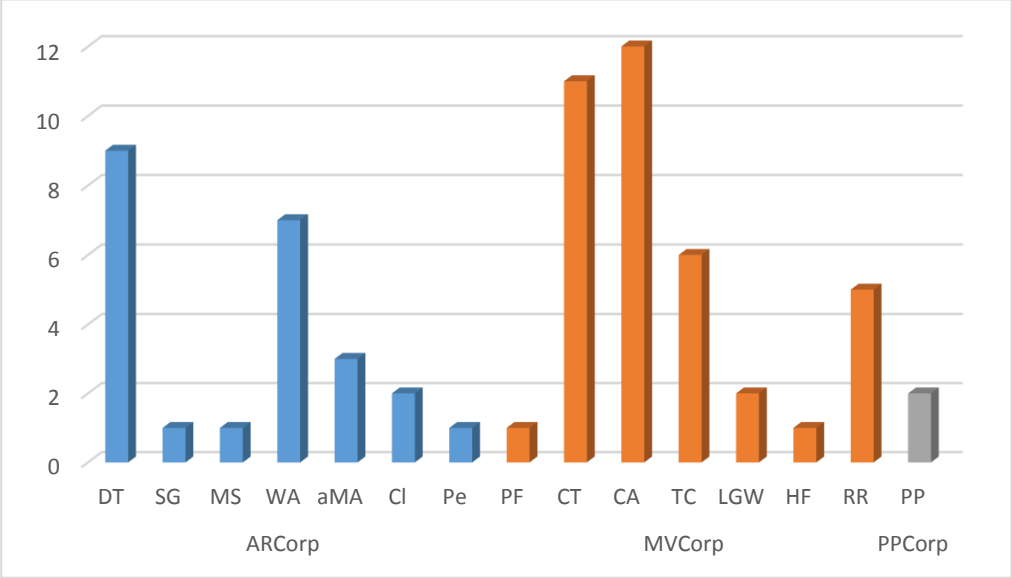
This uneven distribution of batch selection, and the resulting amount of annotation produced by it, is predominantly based on the text size of the individual poems in *MVCorp*. For instance, *Confessio Amantis* (*CA*) and *The Canterbury Tales* (*CT*), the texts demonstrating the highest saturation of modernised forms in Table 7.16, represent two of the longest texts used in this study with token values of 133,985 and 207,382 respectively.[45] The comparatively large sizes of these poems could well have increased the statistical likelihood that batches from these texts would be selected for manual training; this would have resulted in greater amounts of manual training on the orthographic conventions characterising these texts, and ultimately made it easier for the software to subsequently normalise the language in these specific poems during automated annotation.

The correlation between text size and degree of regularised output is also visible when looking at the break-down of source-texts constituting these batches. As can be seen below in Figure 3.14, the textual dispersion produced in the first manual training stage reveals the inadvertent favouring of longer narrative poems over shorter

[45] Refer back to Tables 3.4 and 3.5 for the word counts for each source-text in the reference corpus.

texts. Larger poems in the reference corpus including *CT*, *CA*, *The Destruction of Troy* (*DT*), and *The Wars of Alexander* (*WA*) are shown to have been selected multiple times; when combined, these texts collectively account for 39 out of the 64 (just over 60%) of the total batches annotated in this first stage. This directly proportional relationship between text size and manual training is also exemplified with the shorter poems in *MEPoetLitCorp*, with texts under 20,000 words such as *Mum and the Sothsegger* (*MS*), *Cleanness* (*Cl*), *Pearl* (*Pe*), *The Parlement of Foules* (*PF*), and *The House of Fame* (*HF*) having only been selected once or twice for training for each text.

Figure 3.14: Break-down of texts annotated in first stage of manual analysis from *ARCorp, MVCorp,* and *PPCorp*; data measured by number of 250-word batches (y-axis).



In addition to text size, another factor potentially impacting the saturation of normalised word forms in the post-annotated corpora is the amount of direct manual training conducted for each poem. For example *Joseph of Arimathie* (*JA*), which at 6,384 words constitutes a comparatively small text in *ARCorp*, did not receive any direct manual training on the individual orthographic features characteristic of (and to) that particular text (as demonstrated via its absence from Figure 3.14 above). It is not

difficult to envision how the lack of direct training would produce insufficiently low confidence threshholds for automated annotation to occur, which would offer a reason for the high hybridity of variant and normalised forms present in the annotated excerpt for *JA* revealed back in Table 3.12.

The lack of direct training in *VARD* on *JA* does not represent an isolated occurrence but is rather reflective of a larger trend evidenced in the shorter corpus materials in *MEPoetLitCorp*. This pattern is particularly visible in *ARCorp* including titles such as *The Awntyrs of Arthure* (*AoA*)*, Gologras and Gawain* (*GG*)*, The Parlement of the Thre Ages* (*PTA*)*, Patience* (*Pa*)*, Pierce the Ploughman's Crede* (*PPC*)*, The Pistel of Swete Susan* (*PSS*)*, Richard the Redeless* (*RtR*)*, Saint Erkenwald* (*SE*)*, The Siege of Jerusalem* (*SJ*)*,* and *Wynnere and Wastoure* (*WW)*. In looking at the word counts for these texts provided in Table 3.4, most of these poems are fairly short at 2500-7500 tokens, with the outliers *GG* and *SJ* only containing 10,482 and 11,548 words each. Similarly, the *MVCorp* texts not processed manually including *Anelida and Arcite* (*AaA*)*, The Book of the Duchess* (*BD*)*,* and *The Plowman's Tale* (*PT*) constitute the shortest texts in the corpus with *AaA* at 2,790 words, *BD* at 8,628, and *PT* with 8,827; these numbers stand in sharp contrast to the hundreds of thousands of words in both *CA* and *CT*.

Dialect also constitutes a key factor in the amount of modernised forms observed in the post-annotated source-texts. As initially signposted in 2.4.1, the regional dialects utilised by the late 14th/early 15th century language communities from which these poetic texts originated is reflected in the different phonetically-based orthographic conventions appearing in these poetic texts. Out of the passages illustrated in Table 3.12, those originating from texts written in Chancery English evidenced the highest amounts of regularisation (*CT* and *CA*). Chancery English

constituted an emerging prestige dialect in London influenced by the speech of the East Midlands and the surrounding regions which, with later standardisation, operated as the basis for spelling conventions in E-ME and ModE (Fisher 1977). To this end, texts that were composed in this variant of ME potentially possess a linguistic advantage over the more regional texts present in the corpora; while individualised spelling forms may have changed significantly over time, it would be highly surprising if this historical link between Chancery English and ModE did not factor into the results of orthographic annotation on the basis of comparative linguistic similarity.

By contrast to the Chancery texts, source-texts characterised by strong regional dialects did not respond well to attempts at modernisation. *Sir Gawain and the Green Knight* (*SG*), like the other poems attributed to the Pearl poet, is composed in a ME dialect localised to north-western Staffordshire and south-eastern Cheshire as the presumed variety of language used by both scribe and author (Duggan 2007: 240-1). Texts like *WA* and the alliterative *Morte Arthure* (*aMA*) are similarly regional in containing more northern ME dialects believed to have been used in Northumbria and the north Midlands, respectively. Representative batches for these non-Chancery English texts exemplified in Table 3.12 (*SG* and *WA*) demonstrate notably less uniform regularisation than the batches for *CA* and *CT*, and to this end more ME orthographic variation in the annotated corpus results. In the case of *WA*, the fairly high amount of direct training evidenced in Figure 2.14 as a comparatively long poem does not appear to have drastically overcome the normalisation issues posed by the regional variety of language being used within the text.

To this end, the amounts of modernisation present in the post-annotated corpora appear to be based on a complex network of (at times, competing) factors involving comparative text size, the dialect in which each source-text was composed,

and the amount of contact with *VARD* achieved in the manual training stages. As with the other efficiency measures, this does not represent an exclusive listing of the variables responsible for the different amounts of regularisation evidenced within the annotated corpora but rather suggests the most prominent factors likely to have contributed to the production of the existing corpus data. This information is useful for future corpus approaches orthographically regularising ME text in demonstrating some of the inadvertent tendencies that occur when using unbalanced corpora in terms of genre and overall length, and the approximate amounts of modernisation achieved as a combination of the chosen approach (32,000 manual training) alongside the given texts within each corpus.

3.3.2 Conclusions

While containing some residual issues, the normalised versions of *PPCorp* and *MEPoetLitCorp* overall exhibited sufficiently increased orthographic regularity to be of use for corpus-based analysis. As highlighted in 2.4.1, this usefulness is linked to the correlation between spelling regularity and the quality of the statistically-based quantitative measures underlying corpus analysis, with less spelling variation signalling the likelihood of more accurate corpus results (Baron et al 2009a: 41, Baron & Rayson 2009: 1-2). To this end, my decision to use the normalised versions of *PPCorp* and *MEPoetLitCorp* is underscored by the belief that these versions are capable of generating more accurate data than otherwise possible using the original ME text (Lehto 2010: 289).

Once downloaded from *VARD*, these normalised versions of *PPCorp* and *MEPoetLitCorp* were used to compile the frequency and keyword data on which my corpus stylistic analysis is based. [46]

3.4 Data analysis

The thesis' analysis and discussion chapters have been structured to directly link back to the aims outlined in the study's research questions (1.2). Table 3.15 below sets out each step of the present study's analysis alongside the research question to which each stage of the analysis corresponds.

Table 3.15: Methodology taken in data analysis of the language in *PP*

Step 1	Generate word frequencies and KW data	Answers RQ1 (i)
Step 2	Identify lexical features related to themes of law/consumption	Answers RQ1 (ii)
Step 3	Analyse word patterns (incl. collocation) in each theme to analyse stylistic significance	Answers RQ1 (iii)
Step 4	Compare corpus analysis to literary criticism	Answers RQ1 (iv)
Step 5	Review corpus annotation to discuss effectiveness of orthographic normalisation in corpus analysis of *PP*/ME literature	Answers RQ2 (v-vii)

Within this approach to the language in *PP* Steps 1-2 are addressed in the survey chapter (4), with Steps 3-5 appearing in the main analysis chapters (5 and 6).

[46] Once the corpus data had been generated, the non-annotated versions of the corpora were then used to support the further analysis conducted in Chapters 5 and 6. This decision was made in attempts to remove the possibility of inadvertently replicating accuracy-based misrepresentations produced as a result of corpus regularisation.

3.4.1 Word frequencies and KWs

I first generated word lists for *PPCorp* and *MEPoetLitCorp* in *WordSmith Tools* using the annotated versions of each corpus. These word lists contained lexical entries for all the types and their overall frequency of occurrence in each corpus, with all normalised variants collated into modernised entries. An initial comparison of the frequencies in these word lists was conducted to identify some of the lexical register features attributable to the language in late ME narrative poetry. This was achieved by comparing the top most frequent words in these word lists for *PPCorp* and *MEPoetLitCorp* (see Table 3.16 below) to the most frequent terms found in the *Early English Books Online* (*EEBO*) corpus, a general reference corpus of ME/EModE historical language spanning across a variety of text-types. I also compared the results generated for *PPCorp* and *MEPoetLitCorp* to explore some of the ways in which *PP* departs from the findings in the reference corpus; this comparison focused specifically on the most frequent words generated for each (sub)corpus (see Table 4.2).

Table 3.16: Top 100 most frequent word forms identified out of the 9,250 total in *PPCorp* and 47,928 total in *MEPoetLitCorp*, as generated in *WordSmith*.

	PPCorp		*MEPoetLitCorp*	
N	**Word**	**Freq.**	**Word**	**Freq.**
1	AND	4827	AND	36664
2	THE	2221	THE	35990
3	TO	1767	OF	24679
4	THAT	1762	THAT	24149
5	OF	1530	TO	21408
6	FOR	1169	HE	15862
7	IN	1110	IN	15469
8	Y	1032	HIS	14062
9	A	903	A	13340
10	HE	798	FOR	12854
11	HIS	744	I	12551
12	AS	688	WITH	9816
13	IS	665	WAS	8962
14	HERE	631	AS	8564
15	HIT	595	IS	7684
16	SAID	507	HIM	7669
17	WITH	499	THIS	7043
18	ALL	485	SO	6660
19	NOR	469	IT	6291
20	HIM	467	BUT	6039
21	BE	449	BE	5604
22	HEM	427	THEY	5574
23	ME	380	HER	5161
24	NOT	378	MY	5055
25	MEN	359	ON	5041
26	MY	346	ALL	4814
27	NO	343	NOT	4349
28	YOU	340	AL	4147
29	SO	339	SHE	4021
30	BY	333	WHICH	3954
31	THEY	330	NO	3876
32	BOTH	327	FULL	3860
33	THIS	326	ME	3818
34	HAVE	325	THERE	3810
35	OR	318	NOR	3303
36	THEN	313	HAVE	3285
37	WAS	310	WHEN	3199
38	ON	273	WERE	3176
39	OUR	270	SHALL	3091
40	SHALL	254	SAID	3081
41	AC	253	MAY	2946
42	AL	242	YOU	2941
43	WILL	241	AT	2935
44	WERE	239	WILL	2911
45	AT	232	WELL	2798
46	GOD	228	THOU	2777

47	THERE	224	HAD	2776	75	GOOD	135	THY	1622
48	MAN	220	BY	2755	76	WHEN	132	SHOULD	1615
49	IF	210	THEN	2737	77	LAW	131	HEART	1602
50	OTHER	201	MAN	2706	78	BEN	130	MANY	1595
51	KIND	192	OR	2596	79	UNTIL	127	GREAT	1530
52	TRUTH	190	MEN	2445	80	MEDE	124	T	1407
53	SHOULD	172	HEM	2437	81	LIFE	123	GOOD	1390
54	WELL	172	NOW	2428	82	RIGHT	119	THAN	1357
55	HAD	170	IF	2383	83	ANY	118	SON	1345
56	THO	170	YOUR	2374	84	WIT	118	MORE	1325
57	THY	170	HATH	2299	85	LEAVE	115	RIGHT	1312
58	HOLY	167	OTHER	2273	86	GRACE	114	HIRE	1294
59	MANY	164	KING	2264	87	CHRIST	113	SAY	1291
60	WHAT	163	WOULD	2113	88	VS	111	DAY	1286
61	NEVER	162	YE	2049	89	HATH	109	WE	1272
62	LOVE	161	ALSO	2039	90	KING	109	BEN	1269
63	AFTER	158	WHAT	2031	91	PIERS	109	EVER	1250
64	THROUGH	153	UPON	2012	92	REASON	109	FORTH	1243
65	CONSCIENCE	152	SUCH	1998	93	MORE	108	TIME	1225
66	E	151	LOVE	1941	94	COME	106	FRO	1210
67	MAY	150	THUS	1872	95	CETERA	102	AFTER	1206
68	WOULD	150	GOD	1846	96	THUS	101	NEVER	1201
69	ARE	145	OUR	1842	97	TIME	100	YET	1195
70	BUT	145	MIGHT	1775	98	KNOW	99	AN	1188
71	MIGHT	143	HOW	1756	99	SOME	99	WISE	1153
72	WE	139	OUT	1743	100	DO	98	THEE	1133
73	SUCH	138	HIT	1737					
74	LORD	136	HERE	1636					

These word lists were then used produce a keyword list for *PP* by comparing the poem's language to the text in my reference corpus using *WordSmith*'s KeyWords function. The KW list was created using the entire annotated reference corpus (amalgamated from *ARCorp* and *MVCorp* combined) and to this end incorporated all 76 text-files for *MEPoetLitCorp*. This process identified the 855 total items deemed statistically significant to *PP* based on their unusual recurrence (or textual absence where otherwise expected) relative to the language used in the reference corpus.

This KW list was generated using a p-value of .0001. Using this p-value signifies that these results are 99.99% based on non-coincidental appearance, making any keyword with a statistical significance of 15.13 or greater of unusual importance to (and within) the text.[47] In other words, the keyness attributed to these terms is unlikely to have been based on chance but is rather reflective of the different underlying linguistic patterns contributing to differences in register-based or stylistic form and presentation characteristic of the texts contained within these corpora.

This approach also used log-likelihood as the measure of statistical significance distinguishing these KWs from the rest of the language appearing in *PP*. As the default measurement for *WordSmith*, log-likelihood generates significance (i.e. keyness) by comparing word frequencies in the target corpus to their occurrence in the reference corpus on a case-by-case basis. Log-likelihood is useful for stylistic analysis in that it provides an accurate measure of how 'surprising' the occurrence of a word is, even when occurring in a single textual instance. Kilgarriff (2001: 104) observes that log-likelihood reveals the significance of low-occurrence language events not identified by either mutual information (MI) or χ^2 test, acting not only as an accurate measure of

[47] See Rayson's online LL and size effect calculator (http://ucrel.lancs.ac.uk/llwizard.html; accessed February 2016).

non-intuitive significance in 'finding the most surprisingly frequent words in a corpus without looking at internal structure' but to this end corresponding 'reasonably well to human judgements of distinctiveness'.

Within these set parameters, a brief survey was conducted to identify the leading themes from amongst *PPCorp*'s key lexis (4.2). This approach in exploring the text's themes mirrors traditional semantic domain analysis as a type of early-stage corpus analysis, but the decision to focus on theme is based on the non-separable semantic content of many of these KWs when used in a historical context. Major semantic domains such as religion and politics, for instance, are traditionally categorised separately under the UCREL Semantic Analysis System (USAS) classification system as S9 and G1 respectively. However, these otherwise separate domains are actually unified within the single individual theme of governance in late medieval English culture, wherein religious tenets influenced politics to a degree that is underwritten in the very fabric of social views, beliefs, and customs (Brown 1989). To this end, the thematic classification of the leading concepts emerging from *PPCorp*'s KWs still reflects the aims of semantic analysis in seeking to group together 'word senses that are related by virtue of their being connected at some level of generality with the same mental concept' (Archer et al. 2002: 1), but in a manner that takes into account the historical and literary context surrounding the usage of these key terms.

Within this classification according to lexically-manifested theme, focus was predominantly placed on content words over function words. This decision is based on the observed function of content words in being markers of style through their conveyance of thematic elements central to (and characteristic of) literary texts (Mahlberg & McIntyre 2011: 208-211). While not meaning to imply that function words can never be used as stylistic features in a literary text (e.g. Scott 2000), looking

at grammatical keywords within the structure of the present investigation could inadvertently shift focus to textual indicators of poetic form (i.e., alliterative poetry) over stylistic patternings. This survey additionally focused predominantly on the positive KWs generated for *PPCorp* (see Table 3.11 below) to orient focus on the text's prominently *present* themes as opposed to ones that were unusually absent where otherwise expected based on the content reflected in the reference corpus texts.

Table 3.17: Top 100 lexical keywords out of the 855 total generated for *PPCorp* in *WordSmith.*

N	Keyword	N	Keyword	N	Keyword
1	CONSCIENCE	35	MOUTH	69	ARBITRIUM
2	PIERS	36	JEWS	70	SUNT
3	FORTH	37	EVE	71	TAUGHT
4	MEDE	38	DEI	72	GROW
5	DOWEL	39	SINS	73	GRACE
6	HOLY	40	SOUL	74	AF
7	CHURCH	41	SAITH	75	PRIESTS
8	PORE	42	MEN	76	YES
9	THREE	43	SAVED	77	TRINITY
10	KIND	44	FOOD	78	LIVE
11	SETHE	45	CATTLE	79	LIBERUM
12	CHRISTIAN	46	PARDON	80	MINSTRELS
13	JESUS	47	DOBEST	81	PATER
14	SAID	48	PEES	82	LUCIFER
15	IGNORANT	49	BEGGARS	83	VIRTUES
16	COMMUNITY	50	DRINK	84	LEAVE
17	PLOUGHMAN	51	LOYAL	85	SUFFER
18	COURTESY	52	DEVIL	86	ADAM
19	POVERTY	53	TRUTHFULLY	87	REPENTANCE
20	LAW	54	PEOPLE	88	SOFFREDE
21	TRUTH	55	FRIARS	89	CHRISTENDOM
22	CHRIST	56	WICKED	90	IMPERFECT
23	PRAISE	57	PENANCES	91	BONUM
24	CLERGY	58	CHARITY	92	PSALMS
25	REASON	59	BEARS	93	LIVING
26	THOUGHT	60	GIVES	94	LIVED
27	WITNESSES	61	SIN	95	LIAR
28	DOBET	62	TOGETHER	96	ALE
29	BELIEVE	63	WRIT	97	LIFETIME
30	SPIRITUS	64	HUNGER	98	SULLY
31	TEACHES	65	UNDERSTANDING	99	SINE
32	MERCY	66	UNITE	100	FRIAR
33	CLERKS	67	GUILE		
34	PATIENCE	68	DEUS		

These themes, and the KWs comprising said themes, were identified via initial concordance analysis of the poem's lexical KWs. These concordances provided the means for identifying less salient themes associated with each KW, and also based classification on the majority of textual uses. In this context, a majority is defined as having over 50% of a KW's concordance lines linking the use of the term to a particular thematic concept.

From amongst the identified thematic topics in *PP*, the present investigation focuses on the themes of dietary consumption and the law. In addition to being subjects addressed by existing research in the poem's critical scholarship (2.1.6) these themes collectively encompass large numbers of *PPCorp*'s key content words, albeit greater amounts of legal lexis compared to the consumption-based terms. Selecting themes already featuring in the literary criticism enables the results emerging from the present corpus-assisted stylistic analysis to demonstrate how a corpus approach can contribute to already existing literary research by paralleling ongoing critical discussions addressing similar topical content.

While potentially recursive to methodologically limit focus to themes addressed by the literary scholarship, this approach of focusing on the intersection of existing research acts as an important safeguard in over-interpretation of the generated corpus data. It is additionally worth noting the sheer unlikelihood that a corpus approach would uncover any major textual themes that had not already been signposted in the 130 years of modern Langland scholarship produced since Skeat (1888). Evidence of new information highlighted by corpus methods is far more likely to occur in instances of close analysis with individually selected KWs, not in the general identification of the text's leading themes.

3.4.2 KW selection

After thematic selection in Chapter 4, the analysis chapters (5 and 6) shift focus to – and within – the poem's legal and consumption-based KWs. To start the stylistic analysis, each of these KW groupings were categorised according to Mahlberg & McIntyre's (2011) corpus stylistic framework separating these lexical KWs into fictional world elements and thematic signals. This process enabled selection from amongst the poem's statistically significant KWs for terms of high stylistic potential from a functional perspective. This potential was primarily manifested in grammatically ambiguous word forms and semantically polysemous terms likely to yield rich stylistic patternings in their individual use, as well as to point towards patterns occurring in their themes more broadly. This approach is typically used to classify the KWs generated for an entire text, be it Fleming's *Casino Royale* (Mahlberg & McIntyre 2011) or Conrad's *Heart of Darkness* (Mastropierro 2017); however, my use of this framework aligns with the overarching focus on themes of consumption and legality in the text (ii in RQ1), and to this end produces groupings reflective of the functional approaches taken in the representation of these themes within *PP*.

From amongst these results, my analysis specifically focuses on words bridging these text functions, or 'split-words' as markers of style within the text. Focusing on split-category KWs follows Evans (2018) in identifying such intersections as the linguistic signals of greatest stylistic potential - and thus, interest. Her discussion of 'messy signals' in the works of Aphra Behn, for example, demonstrates how features that are difficult to neatly categorise or else exist as data anomalies often produce the most interesting findings when conducting computational stylometric research. This

concept is no less applicable to corpus stylistic analysis as evidenced, among others, in Matropierro's (2017: 99) discussion of the presence of 'atmosphere' KWs in *Heart of Darkness*, with the uniqueness of their dual functionality as fictional world and thematic signals underlying a characteristic tonal feature of the novel.

The split-words examined in the current investigation include *hunger, glutton,* and *drink* for the consumption-based KWs and *doctor, justice,* and *reason* for the poem's legal KWs. *Hunger, glutton, doctor*, and *justice* are of stylistic interest based on their dual functionality as world-building elements and thematic signals. By contrast, *drink* and *reason* are significant based on their existence as examples fulfilling multiple fictional world functions simultaneously.

3.4.3 Collocational analysis

Analysis of these split-words was subsequently informed by corpus-generated collocations for each of these 6 KWs. These collocations were generated in *WordSmith*'s Concordance function, with collocates defined as words appearing within a window span of L3 to R3 shown to appear in proximity to the node word (i.e., KW) in at least a single textual instance. This 3:3 span enabled a sufficiently broad textual window within which to observe patterns of stylistic interest associated with these KWs, while the lower number of proximal occurrences than those usually set for corpus analysis ensured that no patterns went unobserved in the resulting analysis on account of residual spelling variation in the corpus data.

These collocational patterns were measured for statistical significance according to mutual information (hereafter, MI). While the t-test would work as a measurement of collocational significance based on the relatively small size of my

corpora (McEnery et.al 2006: 55-7), MI was chosen as the pre-set collocational measure in *WordSmith*. The statistical measure also prioritises the significance of the collocation between content words over grammatical ones, which corresponds to the focus on content words addressed in initially generating keywords for *PPCorp*.

As with the KW analysis, this step makes use of the annotated versions of my corpora. This decision was meant to utilise the benefits provided by spelling regularisation in identifying collocational patterns. To this end, each collocational pattern associated with *PP*'s legal and consumption-based KWs are checked against the unannotated text to ensure that the provided collocation does not inadvertently result in patterns not evidenced in the original text (as set out in 3.3.2). In this capacity, any interpretations arising from this collocational analysis are initially identified from the normalised text but are ultimately grounded in the text found in the unannotated *X* manuscript.

Collocational analysis was further supplemented through the use of linguistic theories chosen specifically to align with the given thematic topics. Within the remit of the 6 KWs selected after functional classification, *hunger, glutton, doctor,* and *justice* were analysed using the cognitive and linguistic animacy scales provided by Langacker (1991), Stockwell, (2009), and Foley and van Valin (1985). Animacy scales offer insight into the range of identifiable personhood attributable to individual textual occurrences of each KW, which is significant in revealing the stylistic patterns underlying the production of meaning in each of these multifunctional terms. The other KWs, *drink* and *reason*, were analysed via other linguistic frameworks more aligned to the given thematic topic to which they belong in the poem. *Reason*, for example, was surveyed through the lens of technicality, a correlation highlighting the cline of technicality linked to legal language as a professional language variety; to this end, this

investigation makes use of the legal technicality scale created by Marin (2014) and Marin & Rhea (2014). By contrast, terminology linked to consumption does not operate on the basis of technicality but rather evaluation as an inherently evaluative concept typically associated with positive and negative perceptions of behaviour and general practices of individual and social expenditure. This topical shift thus makes it more useful to examine the consumption-based KW *drink* through a linguistic framework of appraisal (e.g. Martin & White 2005), with evaluation constituting a potentially rich dimension of stylistic significance within this theme.

Studying the poem's KWs through these linguistic frameworks thus enabled observation of repeated stylistic patterns characterising these KWs in *PP*. These patterns demonstrated close ties with overarching features connected to its poetic form (i.e., alliteration and allegory) as well as more KW-specific trends. Connections were also drawn between groups of KWs operating within these themes, particularly those shown to be exhibiting similar patterns of usage in the text. Where relevant, efforts were made to link these observed language patterns to points made in the critical literature.

To conduct this analysis, concordance lines were generated for each of the selected KWs. These concordances were generated from the annotated *PPCorp* but were again cross-referenced against the unannotated text. This was particularly necessary in instances wherein automated annotation produced textual features not present in the original version of the poem.

In the course of conducting this collocational and concordance-based analysis, corresponding data was drawn from *MEPoetLitCorp*. This additional information was used for comparative purposes in demonstrating how the use of these KWs and their collocates in *PPCorp* are similar to – or deviate from – the lexical patterns evidenced in

the study's other late medieval poetic texts. This provided a broader literary context in which to situate the observed patterns within *PP*, thus presenting a basis for explorations of exhibited (ab)normality in their textual use. While the legal analysis section made use of the entire *MEPoetLitCorp* holistically, it was deemed useful to break the reference corpus into its sub-corpora (*MVCorp* and *AVCorp*) to illustrate how poetic form played a role in the types of relationships evidenced in the consumption-based KWs.

3.4.4 Contributions to literary criticism

In the process of conducting the linguistic analysis set out in 3.4.1-3.4.3, the final component of this investigation lies in integrating my corpus-assisted findings with ongoing discussions on the language of dietary consumption and legal language present in Langlandian scholarship. This approach provides the opportunity to identify the value of my corpus data and analysis (and by extension, a corpus approach to the poem more generally) as results that either corroborate existing findings, build on existing observations and theoretical frameworks, or offer new insights into *PP*'s language that have not been previously addressed. These observations are interspersed throughout the analysis chapters (4-6), with a general summation provided in the conclusion (Chapter 7).

4. SURVEYING THE CORPUS: FREQUENCIES AND KEYWORDS IN *PPCORP*

This introductory analysis chapter surveys the initial corpus results produced from the comparison of *PPCorp* to *MEPoetLitCorp* using *WordSmith Tools*. 4.1 compares the generated word lists and top frequencies for both corpora to signpost some of the register features evidenced across late medieval poetic text types, as well as features revealing differences between the separate literary traditions. 4.2 focuses on the keyword results generated for *PPCorp* and categorises them into the semantic domains rendered prominent by the KW data. This classification reveals the text's salient themes (and the keywords within each theme), with discussion placing particular emphasis on the KWs contributing to the themes of law and consumption explored in greater detail in the forthcoming chapters (5 and 6).

4.1 Word frequencies

As mentioned briefly in 2.3.3, word frequencies measure the raw frequency by which each word occurs in a given corpus. These measures appear in word lists generated by counting of all the words within a corpus, wherein the total *types* (distinctive word forms in a text or corpus) are ordered by descending frequency of textual appearance.

Given the ability of frequency measures to reveal characteristics of a text or a certain variety of language (Scott 2001: 47), examining these word lists offers a view by which to explore the language use in *PP* relative to the other medieval poems contained in the reference corpus. This comparison was effectuated by generating word lists for both *PPCorp* and *MEPoetLitCorp*, with Table 4.1 below providing the top 20 most frequently occurring words in each corpus as an arbitrary cut-off point. A selection containing the top most frequent words from the *Early English Books Online*

(hereafter, *EEBO*) corpus[48] is also included in Table 4.1 to illustrate the types of high-frequency words found in general language use within a historical context in order to better contextualise the information being generated for my specialised literary corpora.[49] These highly frequent words are listed in descending order of raw frequency, with the 'texts' column signifying the number of text-files in which each word appears.

As demonstrated below in Table 4.1, the top sections for all three corpora are dominated by closed-class words. This high density of grammatical words is an unsurprising finding that correlates with the general expectation in corpus linguistics that grammatical terms will constitute the most frequent types in a corpus as the structural elements holding together the discourse (Scott & Tribble 2006: 15-6) in virtually all language varieties (McEnery et al. 2006: 53). The results exhibited in Table 4.1 prominently feature common grammatical words including definite and indefinite articles (*the, that, this, a*), prepositional elements (*to, for, of, in, with, by*), conjunctions (*and, nor, so, but, as*), auxiliaries (*is, be, was*), and a range of demonstrative, relative, and personal pronominal forms (*this, that, y* 'I', *I, he, his, him, here* 'her/their'*, hit* 'it', *it, they*). The ordering of these closed-class words appear to broadly correspond with one another with only slight differences in ranking most likely linked to the size differences between the respective corpora (*EEBO* versus *PPcorp* and *MEPoetLitCorp*).

[48] Open access to the word list for EEBO available through the *Early Modern OCR Project* website (http://emop.tamu.edu/eebo-word-freq?items_per_page=50).
[49] It is worth noting here that the 1470-1709 range covered by *EEBO* only marginally overlaps the current 1350-1500 time span covered in *MEPoetLitCorp* making it more suitable as a general language corpus for a study of EME rather than ME, but still remains more fit for purpose than a modern general language corpus like the *British National Corpus* (*BNC*).

Table 4.1: The top 20 most frequent words generated for *PPCorp* (left), *MEPoetLitCorp* (centre), and *EEBO* (right).

PPCorp

N	Word	Freq.	Texts
1	AND	4827	1
2	THE	2221	1
3	TO	1767	1
4	THAT	1762	1
5	OF	1530	1
6	FOR	1169	1
7	IN	1110	1
8	Y	1032	1
9	A	903	1
10	HE	798	1
11	HIS	744	1
12	AS	688	1
13	IS	665	1
14	HERE	631	1
15	HIT	595	1
16	SAID	507	1
17	WITH	499	1
18	ALL	485	1
19	NOR	469	1
20	HIM	467	1

MEPoetLitCorp

N	Word	Freq.	Texts
1	AND	36664	75
2	THE	35990	73
3	OF	24679	76
4	THAT	24149	74
5	TO	21408	75
6	HE	15862	75
7	IN	15469	74
8	HIS	14062	73
9	A	13340	76
10	FOR	12854	75
11	I	12551	75
12	WITH	9816	71
13	WAS	8962	65
14	AS	8564	76
15	IS	7684	72
16	HIM	7669	66
17	THIS	7043	75
18	SO	6660	73
19	IT	6291	68
20	BUT	6039	76

EEBO

N	Word	Freq.
1	THE	15,059,595
2	AND	10,110,480
3	OF	9,775,138
4	TO	8,048,287
5	IN	4,993,275
6	THAT	4,643,117
7	A	3,661,493
8	IS	3,159,644
9	IT	2,909,193
10	BE	2,549,071
11	FOR	2,481,585
12	AS	2,384,344
13	HIS	2,307,104
14	NOT	2,165,513
15	BY	2,058,431
16	HE	1,986,495
17	BUT	1,955,116
18	I	1,915,095
19	THEY	1,834,821
20	WITH	1,825,972

From within the grammatical data listed in Table 4.1, the most striking difference across the corpora is found in the presence of the conjunction *and* as the most frequently-occurring word in *PPCorp* and *MEPoetLitCorp*. This sets my specialised corpora apart from most other varieties of English, within which *the* consistently appears as the most frequently used word (e.g. Scott & Tribble 2006: 24). The recurring primacy of *the* is shown in a modern context through Leech et al.'s (2001)[50] work with the *BNC* – a general language reference corpus of modern British English – who observe that *and* consistently appears as one of the most common words in the *BNC* but is never the most frequent word in its sub-categories; it appears as the 3rd most-common word in the spoken component but only as the 4th most frequent term in the written section of the corpus. The appearance of *the* as the most frequent word in *EEBO* as shown in Table 4.1 additionally suggests that the unusually frequent use of *and* demonstrated in my specialised corpora represents less a difference between modern and historical varieties of English than a feature reflective of the medieval poetic text-types contained within my corpora.[51]

This increased use of *and* indicated in Table 4.1 corresponds with existing scholarship identifying grammatical coordination as a register feature characteristic of medieval literary texts. Historical syntactic preferences for parataxis, or the juxtaposition of main clauses over clausal subordination, is regularly signposted as a feature of older English literatures compared to modern literary conventions (Horobin & Smith 2002: 102, van Gelderen 2018: 8, 35-7). This clausal coordination, in addition to the phrasal coordination evidenced within the intra-linear binomial pairings in the alliterative texts (e.g. *life and lyflode*, *right and resoun*), is particularly common in line-

[50] Based on the frequency-based lexicographical work conducted by Kilgarriff (1997).
[51] The presence of *and* as the most common lexical unit occurring in these medieval texts has been corroborated by word lists generated for the non-annotated versions of my corpora.

initial positions as a means of linking related events occurring over multiple poetic lines.[52] Examples of this are visible across both corpora, as can be seen in the description of beggars in *PP* (i) and the death of Crusea alluded to in *Confessio Amantis* (*CA*) (ii).

(i) In glotonye þe gomus goth þei to bedde
And ryseþ with rybaudrye… (*PP*, Prol. 44-5)

In gluttony the men go off to bed
And *rise with ribaldry…*

(ii) Anon therof the fyr sprong oute
And brente hir bothe fleissh and bon. (*CA*)[53]

Soon from [the dress] the fire sprung out
And *burnt both her flesh and bone.*

In addition to the core function of promoting inter-linear cohesion (i and ii), these line-initial uses of *and* also contain additional aesthetic features. These are particularly notable in instances involving structural repetition, as exemplified in *PP*'s discussion of hermits (iii) and also in the listing of superlative qualities attributed to Sir Gawain in the alliterative *Morte Arthure* (*aMA*) (iv).

(iii) Summe hadde lyflode of his lynage and of no lyf elles
And summe lyuede by here letterure and labour of here handes
And somme hadde foreynes to frendes þat hem fode sente
And briddes brouhte **somme** bred þat they by lyuede (*PP*, 9:202-5)

Some had food and drink from his family and no one else
And some *lived by their learning and the labour of their hands*
And some *had foreigners for friends that sent them food*
And to some *birds brought bread that they lived on*

[52] For instance, see the abbreviated *and* (7) appearing line-initial in the first 68 lines in the early ME poem *The Owl and the Nightingale*, as reproduced in van Gelderen (2018: 115).
[53] As with the textual excerpts for *PP*, the passages taken from *MEPoetLitCorp* are derived from the non-orthographically annotated versions of the reference corpus.

(iv) This was Sir Gawain the good, the gladdest of other,
And the graciousest man that under God lived,
Man hardiest of hand, happiest in arms,
And the hendest in hall under heven-rich,
And the lordliest in leding whiles he live might… (*aMA*, 3876-80)

He was Gawain the good, the gladdest
And the most gracious *man that under God lived*
The man hardiest of hand, most happy in battle
And the humblest *in the hall under heaven*
And the lordliest *in leadership as long as he might live…*

The present corpus data appears to confirm the dominance of syndetic parataxis in my sampling of medieval poetry. This dominance is more preferential than exclusive, with Horobin and Smith (2002: 102) noting that an increased number of hypotactic constructs started to emerge in the comparative styles of different ME authors (for example, Chaucer's hypotaxis compared to Malory's parataxis). This divide in clausal structuring trends appears to broadly correspond with differences in poetic tradition, with the greater presence of continued parataxis likely to be found in Alliterative Revival texts like Malory's alliterative *Morte Arthure* (*aMA*) and *PP*. This is significant for *PP* because paratactic constructions 'require readers to make causational links which are not made explicitly by writers, whereas writers who employ hypotaxis make such links explicit' (Horobin & Smith 2002: 103). This link between textual parataxis and high levels of cognitive inference, in addition to the other text features outlined in 2.1.2, could potentially contribute to some of the interpretative difficulties attributed to *PP* as a text that makes greater cognitive demands of its readers (particularly, modern readers) by adhering to a larger discoursal trend characterised by systemic lexico-grammatical coordination.

In addition to closed-class items, the high frequency open-class words generated for *PPCorp* and *MEPoetLitCorp* reveal information about the character and

composition of my corpora. These indicators of content operate alongside a limited number of lexical words clustered within the top 100-ranked words of a given corpus that typically express information 'concerned with humans and what they have recently said, known, and seen' (Scott & Tribble 2006: 24) including words referencing individuals (*man, king, lord, god, son*) and groups of persons (*men*), as well as common actions performed by said persons (*said, say, come, leave, know*).[54] A more focused view of the frequent content words contained in my corpora is provided below in Table 4.2, with the results for *MEPoetLitCorp* broken down into its sub-sections *ARCorp* and *MVCorp* for greater detail.

[54] Examples cited here are drawn from Table 3.16, which lists the top 100 words contained (grammatical and lexical) in *PPCorp* and *MEPoetLitCorp*.

Table 4.2: The top 50 most frequent content words generated for *PPCorp* (left), and *AVCorp* (centre), and MVCorp (right). Terms listed in descending order based on raw frequency of occurrence, with the 'texts' column signifying the number of texts in which each word appears within its respective corpus.

N	Word	Freq.	Texts
16	SAID	507	1
25	MEN	359	1
46	GOD	228	1
48	MAN	220	1
51	KIND	192	1
52	TRUTH	190	1
58	HOLY	167	1
59	MANY	164	1
62	LOVE	161	1
65	CONSCIENCE	152	1
74	LORD	136	1
75	GOOD	135	1
77	LAW	131	1
80	MEDE	124	1
81	LIFE	123	1
82	RIGHT	119	1
83	ANY	118	1
84	WIT	118	1
85	LEAVE	115	1
86	GRACE	114	1
87	CHRIST	113	1
90	KING	109	1
91	PIERS	109	1

N	Word	Freq.	Texts
31	KING	1459	17
39	MEN	1256	18
43	MAN	1142	18
49	SAID	923	18
51	MANY	882	17
53	HOME	851	13
57	GREAT	699	18
64	SIR	629	12
69	SON	584	18
72	LORD	503	18
74	COME	499	17
78	KNIGHTS	491	11
80	TIME	476	17
83	DEATH	461	15
84	FELL	461	17
86	GOD	448	18
88	MORE	443	18
90	HEART	429	18
91	ANY	427	18
92	BATTLE	424	16
94	FAIR	417	17
98	LAND	405	17
99	DAY	403	18

N	Word	Freq.	Texts
39	SAID	2158	55
45	LOVE	1791	36
58	MAN	1564	51
60	GOD	1398	51
67	MEN	1189	49
68	HEART	1173	39
70	RIGHT	1159	43
75	SAY	1047	50
77	GOOD	1027	45
79	THING	997	41
82	BEGAN	950	40
86	TELL	884	52
87	DAY	883	39
88	MORE	882	43
89	FORTH	852	39
90	GREAT	831	37
91	WISE	828	37
92	MAKE	823	43
94	KING	805	27
95	TAKE	782	37
98	SON	761	32
99	SEE	753	43
100	TIME	749	40

92	REASON	109	1	101	FORTH	391	17	103	KNOW	724	42
93	MORE	108	1	102	HAND	391	18	106	MANY	713	41
94	COME	106	1	103	SET	386	18	112	LET	672	45
97	TIME	100	1	104	NOBLE	385	16	113	THOUGHT	651	34
98	KNOW	99	1	107	CLEAN	379	17	115	MADE	645	35
99	SOME	99	1	109	BOLD	371	16	116	SET	638	34
101	MADE	98	1	111	GREEKS	364	3	123	ANY	601	42
103	CHURCH	95	1	112	RICH	364	17	124	SOME	591	41
106	PEOPLE	94	1	113	GOOD	363	18	125	COME	581	38
107	LET	93	1	114	HOR	362	6	128	TOOK	558	34
108	PORE	93	1	117	GROUND	357	17	129	SAW	557	34
109	RICH	93	1	118	HIGH	357	15	133	LORD	516	34
110	MAKE	91	1	121	GOLD	354	17	134	WIFE	516	37
112	SIN	91	1	124	HEAD	345	16	135	DEED	511	35
114	MERCY	88	1	125	FOLK	342	15	137	TRUTH	509	33
115	FORTH	86	1	126	FIRST	340	16	138	CERTAINLY	504	46
117	HELP	84	1	127	WORDS	334	15	141	MOST	498	32
119	LIVE	83	1	128	ARMS	329	12	143	WORLD	493	33
120	CAM	80	1	129	SE	328	17	146	GO	487	34
121	HEAVEN	80	1	130	WISE	325	16	148	TOLD	487	33
122	SAY	80	1	131	LEDE	324	16	149	LADY	486	28
126	SOUL	74	1	132	EARTH	312	13	150	LIFE	486	35
127	GODS	73	1	133	MUCH	310	14	151	DONE	484	41
128	MUCH	72	1	136	MADE	304	17	152	FATHER	471	27
131	IGNORANT	68	1	138	WIT	304	18	155	WO	470	32
132	LORDS	67	1	139	DERE	301	17	156	FIRST	468	40
133	TRUE	66	1	140	LEAVE	300	18	159	BEST	440	35

These commonly-recurring content-specific lexical words illustrated in Table 4.2 do not identify my corpora's texts as being fundamentally literary when viewed in contextual isolation, but do reveal topical differences rendering each corpus distinctive from one another. For instance, *ARCorp* is characterised by high frequency content words indicative of militant feudalism, including key figures and demographics (*knights, sir, lords, men, greeks*), actions involving these individuals (*battle, fell, forth, death*), potential reasons or incentives for said conflict (*ground, land, earth, gold*), and descriptions through which said events, persons, and actions are portrayed (*bold, arms, heart, noble, great*). *MVCorp* similarly utilises some of the same lexis identified for *ARCorp* (e.g. *lord, king, heart*) but appears to do so within the context of developing narratives focused more on romance than warfare; this lexis incorporates a larger variety of persons denoted by social or familial role (*father, lady, wife*) with focus on emotions (*love, heart, wo[e]*) of high intensity or priority (*best, most, first, right*) with possible results involving marriage, abduction, or other related actions (*deed, take/took, make, life*). By contrast, *PPCorp*'s frequent content words reveal a more comparatively religious topical focus featuring recognisable Christian figures and institutions (*god, church, christ, heaven*), some of its salvational aims (*mercy, help*), evaluative descriptions of actions or behaviours in light of these principles (*good, sin, right*), and the things impacted (*life, soul, conscience*) or otherwise achievable by adherence to these type of values (*grace*).

The topical generalisations realised through the data in Table 4.2 are predominantly correlative to the contents of the largest texts found in each corpus. The militancy evidenced in *ARCorp*'s frequent lexical words, for example, can be traced to the combined presence of several Arthurian legends (*aMA, AoA, GG, SG*) appearing

alongside poetic retellings of historical military events (*DT, SJ, WA*), with *The Destruction of Troy* (*DT*), *The Wars of Alexander* (*WA*), and *aMA* constituting the three largest texts in the corpus (see Table 3.4). Similarly, the more romance-oriented content evidenced in *MVCorp*'s highly frequent content words appears to predominantly stem from Gower's *Confessio Amantis* (*CA*), as well as a selection of Chaucer's shorter romances including *Anelida and Arcite* (*AoA*), *The Romance of the Rose*, (*RotR*), *Troilus and Criseyde* (*TC*), and parts of his *Canterbury Tales* (*CT*).

Thus far, this cursory examination of high frequency patterns has provided a useful overview revealing some interesting insights about the language contained in late medieval poetry and topical differences characterising the major texts within the reference corpus, but offers little detailed information on the linguistic characteristics rendering *PP* distinctive from the rest of *MEPoetLitCorp*. The data contained in Table 4.2 points towards religion as a core topic addressed in the poem with items like *conscience* and *soul* suggesting promising places of analytical departure, but further keyword analysis is needed to generate a more manageable set of statistically salient lexical items than is otherwise possible using *PPCorp*'s word list. These KWs are likely to indicate the most comparatively characteristic themes for *PP*.

4.2 Keywords and themes

As discussed in Section 2.3.3, keywords identify words that 'are comparatively more (or less) frequent in a specific text (or corpus) than in a reference corpus' (Mahlberg 2013: 20), wherein the unusual presence of words in the target corpus compared to the reference corpus occurs to a degree that is statistically significant. Categorising and making associations between keywords constitutes an integral step in the current

investigation enabling me to identify lexical sets contributing to themes of law and consumption in *PP*, which provides a point of departure for exploring the ways in which these themes are stylistically conveyed. In this way keyword analysis continues a lexis-driven approach to the text which, through its salient lexical items, has the potential to gain access to aspects of the text that are perhaps obscured, or otherwise not intuitively obvious (Bondi 2010: 5).

Comparing *PPCorp* to *MEPoetLitCorp* using *WordSmith*'s *KeyWords* tool produced a total of 855 keywords for the poem (see Figure 4.3 on the accompanying CD). To narrow these results to more manageable levels for KW classification all function words were deleted, as well as items possessing negative keyness values. The first decision follows studies such as Mastropierro (2017: 90) in arguing that content words are more likely to reflect major themes than grammatical items because they provide more information regarding the 'aboutness' of a text through their conveyed semantic content. Excluding the negative KWs generated in Table 4.3 similarly aligns with the current research's aim to focus on the thematic content expressed with unusual prominence in the poem, as opposed to the *absence* of expected thematic material.

Focusing on *PPCorp*'s positive lexical KWs narrowed thematic analysis to the 637 entries used for KW classification. While a large number of KWs compared to the amounts used to classify KWs in other types of CS analyses (for instance, the 239 KWs used in Mastropierro 2017) a wider range of KWs for initial classification enabled more comprehensive identification of the text's key themes to compensate for potential variability in the generated KW results. This did not mean that 637 distinct words were classified, for many of the KW entries contained duplications caused by residual spelling variation. Examples of these duplications generally involved one form that was

normalised (e.g. *contrition*), then 1-2 other ME variants (*contricioun*) and, at times, the equivalent term in Latin (*contricio*). To classify these KWs I examined the concordances for each lexical item to contextualise the range of thematic categories to which each word contributes, with categorisation ultimately dependent on the theme reflected in the majority of its concordances. In terms appearing with high frequency (for instance, the 190 occurrences of *truth* in *PP*), I used Sinclair's (1999) method of classification via a sampling of 30 randomly-selected concordance lines.

These key themes, as compiled based on the generated KW data, are presented below in Table 4.4. For the purposes of conciseness, key themes containing fewer than 10 KWs (including themes such as 'learning' and 'conflict') have not been included in the below categories. While only a selection of representative KWs are reproduced here, these terms listed in Table 4.4 provide an overview of some of the text's most salient themes as represented via its generated keywords.

Table 4.4: Primary thematic categories of *PPCorp*'s keywords

Themes		**Example keywords**
Religion	Virtues and vices	*sin(s), sloth, guile, loyal*, pride, patience, courtesy, covets, lechery, unkindness, grow*, clothed, syn(ne)fol(e)* 'sinful',
	Salvation	*believe, soul(s), repentance, faith, shrive, save(d), contrition, foregiveness**
	The Church	*holy, christ(ian), devil(s), psalm, trinity, lucifer, adam, bishops, cardinal(s), mass, eden, god, angeles, messiah, david, shep*
Law		*clerks, witness(es), justice, conseyl(e)* 'counsel', *claims, foregiveness*, writ, loyal*, reason, caes, arrears, vsure** 'usury'
Finance		*pore, seluer* 'silver', *money, erne, mede** 'financial compensation', *poverty, vsure*, tithe, rich*
Nature		*summer, fertilise, weathers, creatures, beasts, winter, cultivate, plonte, grow*, kind (wit)*
Consumption		*food, drink, hunger, ale, beg(gars), bred, moderation, friday(s), brewer(s), glutton, fasting*
Labour		*work, labour(ers), wastrels, plough* 'plow', *ploughman, team, mede**
Community		*community* 'commune', *people, patriarchs, neighbor, kin, persons, mankind, brethren(e), fellowship*

*denotes split KWs across several thematic categories

As evidenced above, the lexis revealed in this thematic KW classification expands on the topical content identified via *PPCorp*'s high-frequency words in 4.1. Terms relating to religion have not only been reproduced but also divided into further sub-categories reflecting the text's approach to religion which, as set out by the current KW categories, is shown to be characterised by doctrine-based discussion of

behavioural qualities (virtues and vices) and the salvation possible (or not) based on these behaviours. These key terms appear alongside recurring reference to the Church as an institution via theological symbols (*trinity, devils*), biblical figures (*adam, eve, david*), and real-world professional roles (*bishops, cardinals*).

The KWs in Table 4.4 additionally provide a more lexically comprehensive picture of *PP* revealing a more detailed range of themes addressed in the poem including consumption, nature, labour, finance, community, and the law. The highly political nature of *PP* as a text concerned (and actively engaging with) ongoing events in a late 14th century context (Barr 1993) makes it unsurprising to find themes addressing complex legal and socio-economic dynamics represented in its key lexis. In addition to the legal and consumption-based KWs (which are examined in greater detail in the forthcoming chapters), the finance category is represented via lexis relating to items (*money, seluer*) or processes (*mede, vsure, tithe*) involved in the exchange of material wealth (or in the case of *poverty*, the lack thereof). The text's labour-based key lexis primarily features the types of work (*work, labour*) and persons conventionally performing said labour (*labourers, ploughman*) contrastive to those characterised by not working (*wastrels*), while the community-based KWs are represented by high numbers of common nouns covering demographic groups (*patriarchs*), relationship markers (*kin, neighbor*), or more universal referents (*people, persons, mankind*).

By contrast to the other categories, the nature-based KWs are characterised through a range of verbal processes (*fertilise, cultivate, plonte, grow*) as well as nominal terms identifying seasons (*summer, winter*) and non-human entities (*creatures, beasts*). It also includes the lexical unit *kind wit* which, formed from the co-occurring combination of the KWs *kind* and *wit*, refers to the idea of naturally-derived

reasoning or wisdom. Considered an inborn mental capacity to distinguish between right and wrong (*MED*), *kind wit* stands in opposition to cognitive processes fuelled by either divine revelation or learned types of logical reasoning; White (1988: 3-5) discusses a number of modern equivalents for this concept offered by medievalists, including 'practical' or 'natural reasoning', 'native intelligence', and even 'common sense'.

The categories listed in Table 4.4 are not self-contained units but rather overlap in parts, as signposted via the presence of split-words in multiple thematic categories. For example, *forgiveness* is used in *PP* both in a religious context as a pardoning or absolving process as well as referring to the waiving of a debt or obligation in a legal sense, while *vsure* 'usury' as the lending of money at disproportionate rates of interest has significance as a financial process with some degree of legal oversight. *Mede* similarly operates at the intersection of *PP*'s financial and labour-based themes by generally alluding to fiscal earnings or some other type of material-based compensation (*MED*), but is framed in the text as a type of corruption-based bribery antithetical to honestly acquired and earned wages. The KW *lele*, normalised to 'loyal' in the annotation process, also recurringly operates as both a religious and legal KW through its meanings of loyalty and truthful lawfulness, respectively.

The pragmatic significance embedded in these split KWs thus represents one of the ways in which *PP* uses language to contribute to the interconnectivity of its leading textual themes. This predominantly occurs via polysemous lexis which, in addition to making use of its various senses (e.g. *forgiveness* and *loyal*), also achieves this thematic bridging through further metaphorical usage. The KW *grow*, for example, is frequently used to describe the physical maturation of natural elements such as *weeds* (1), *fruit* (2), *grass* (2), *apples* (1), *wheat* (1), and even *beasts* (1) but also recurringly describes

the more abstracted development of behaviours and entities possessing religious overtones including *pride* (1), *devil* (1), *gluttony* (1), *concupiscencia* 'excessive desires' (1), *holiness* (1), *grace* (1), and *wrath* (1). The figurative combination of these natural and religious themes through the use of this verb results in rather sophisticated imagery, as exemplified in the textual passage below discussing the acceptability of children born out of wedlock (v):

(v) For god saith hym-sulue shal neuere goed appel
Thorw no sotil sciense a on sour stok **growe** (*PP*, 10:210-11)

For God says himself, 'never shall a good apple
*Through no subtle science **grow** on a sour stock'*

This focus on the split-words evidenced in the poem's thematic KWs serves a dual purpose. From a corpus linguistic perspective, these terms draw attention to the fuzzy boundaries between these topics, and some of the methods by which language is used to cross interchangeably between these key themes. The demonstrated penchant for split terms to be used metaphorically in various thematic contexts also corresponds to Mahlberg and McIntyre's (2011: 212-3) observation that lexical outliers revealed during KW classification have the potential to be of stylistic interest. Within *PPCorp*'s KWs, these split-words include not only verbs like *grow* and *cultivate*, but also nominal and adjectival KWs like *mede* and *loyal* as items of potential stylistic interest based on their dual thematic import and metaphorical potential as linguistic approaches likely mirrored in the poem's legal and consumption-based KWs.

4.3 From corpus survey to textual analysis

From amongst the data compiled in Table 4.4, the forthcoming analysis chapters (5 and 6) focus on the textual patterns utilised in framing themes of consumption and the law in *PP*. This decision aligns with the research aims set out in the research questions in 1.2 (iii in RQ1) and is justified not only through the statistical salience evidenced for each of these themes via their KWs, but also through the critical interest in these topics evidenced in the text's literary criticism, as outlined in 2.1.6. To this end, these analysis chapters take the thematic and consumption-based keywords generated for *PPCorp* in Chapter 4 as points of departure for investigating how these key lexical features are stylistically deployed in the poem.

5. LEXICAL REPRESENTATION OF CONSUMPTION

This chapter conducts a detailed examination of the keywords generated for *PP* addressing consumption. Within these KWs I use corpus stylistic frameworks and information from the corpus-generated collocations and concordances to reveal recurring patterns that may have stylistic import in the text. This type of analysis necessarily takes into account the historical views and values concerning consumption, as well as signposting how normative these results are compared to the poetic reference corpus.

5.1 starts by introducing the complete listing of consumption-based keywords. These terms are then categorised into Mahlberg and McIntyre's (2011) CS keyword framework in 5.2, with specific focus on the stylistic potential embedded in the key lexis bridging different functional categories (e.g. fictional world and thematic signals).[55] From within these bridging or 'split' KWs, three are selected for further analysis in the remainder of the chapter: *hunger, glutton,* and *drink*. 5.3 focuses on *hunger* and *glutton* as keywords containing dual text functions as allegorical personifications, while 5.4 examines *drink* as a term operating as two different types of fictional world elements. To this end I use Hoey's (2004, 2005) theory of lexical priming to link these thematic and world-building functions to specific textual patterns in the use of *hunger* and *glutton*, while my analysis of *drink* is supplemented by Martin and White's (2005) linguistic appraisal framework to illustrate the text patterns linked to highly evaluative consumption-based concepts like drink(ing) in medieval literary discourse. These corpus findings are linked back to relevant discussions taking place in the poem's surrounding literary criticism, as detailed in 5.3.4 and 5.4.4.

[55] For a review of the parameters differentiating fictional world elements from thematic signals, see 2.3.3.

5.1 Keywords

Out of the 855 keywords generated for *PPCorp*, the text contains 26 relating to practices of dietary consumption. This KW grouping does not cover all terms contributing to thematic consumption in the poem, merely those identified as being unsusually salient compared to the other reference corpus texts. This remit also focuses only on KWs relating to consumption as a dietary process as outlined in (7a) within the *MED*, as opposed to all aspects of consumption within a late 14th century worldview. Of these eligible terms, only those with over 50% of their concordance lines addressing matters relating to dietary consumption are included in the present investigation.

The KWs representing consumption in *PP* are listed below in Table 5.1. These words are ordered by descending keyness value, with the N-value denoting the keyword's overall ranking in the poem's full KW list. As discussed in 3.3.1, the keyness of these terms to *PP* was calculated through log-likelihood using a p-value of .0001 and a base keyness value of 15.13. Including this range of keyness enables more complete exploration of how consumption is lexically represented in *PP* through highly unusual words, as well as those less statistically significant compared to the other medieval poems in the reference corpus.

Table 5.1: List of the 26 orthographically normalised KW forms generated for *PPCorp* representing dietary aspects of consumption

N-value	Keyword	Keyness value
67	*food*	99.80
73	*drink*	97.59
90	*hunger*	75.26
133	*ale*	52.60
236	*beg*	35.20
280	*starving*	31.32
301	*fruit*	27.89
367	*lene*	25.60
382	*glutton*	23.55
384	*friday*	23.40
394	*eat*	22.62
400	*lent*	21.61
401	*mete*	21.49
407	*bacon*	21.00
408	*bred*	20.95
410	*brewer*	20.88
412	*brewers*	20.88
415	*moderation*	20.88
421	*fridays*	20.88
480	*flesh*	20.86
498	*begs*	20.84
526	*fasting*	18.20
545	*begging*	17.03
613	*temperancie*	15.66
626	*bakers*	15.66
706	*y-dronke*	15.66

As shown above, *food* (67) and *drink* (73) appear as the most significant of the consumption keywords with keyness values of 99.80 and 97.59. The high statistical significance attributed to these terms reflects their topical relevance to matters of consumption in the poem, with the similarity evidenced in their keyness values to some extent derived from the regularisation of the collective Middle English term *lyflode* into its separate modernised components of ‘food’ and ‘drink’. Other terms appearing near the top of Table 5.1 feature the consumable product *ale* (113), as well as the name of the allegorical character *hunger* (90) in representing a state of dietary non-consumption. In contrast to this handful of highly key terms, the majority of words

featured in Table 5.1 are shown to possess keyness values under 30. This part of the KW listing prominently features terms relating to specific types of food including *fruit* (301), *mete* 'meat' (401), *bacon* (407), *bred* 'bread' (408), *flesh* as a synonym for meat (480), as well as the persons who supply these products such as *brewer*(*s*) (410, 412) and *bakers* (626). The least key entries appear to largely address actions relating to a lack of consumption through *fasting* (526), or instances of attempted consumption through *begs* (498) and *begging* (545).

When finalising this keyword list, a number of terms appeared to contribute to representations of dietary consumption but were ultimately not chosen for further analysis. These were omitted based on initial concordance analyses of the non-annotated text, which revealed the majority of their uses in the poem (over 50% of its concordances) to be unrelated to consumption. The keywords *mouth* (58) and *tongue* (375) for instance, conceivably body parts linked to consumption, are almost exclusively used in *PP* in reference to dialogue, with the tongue portrayed as the tool (and the mouth the receptacle) through which speech is articulated. Another omitted keyword was *shep* (354); while the animal does tangentially relate to dietary consumption as an edible consumable not unlike *mete* or *bacon*, it is generally specific *parts* of the animal that are consumed, not the entire animal.[56] This distinction serves as the reasoning for excluding other keywords in *PPCorp* such as *cattle* (68) and *goats* (565). Concordance analysis furthermore assisted in removing from consideration homonymous terms such as *mede* (17), which can be used to refer to 'mead' as the beverage – be it alcoholic or drink in general – as well as 'meed' as a gift, endowment, or similar type of financial award. Given its single textual use to refer to the drink

[56] In addition, *shep* is typically used in *PP* to metaphorically refer to religious followers in a collective sense.

compared to the 124 total occurrences of *mede* as financial reward in *PPCorp*, its presence as the most statistically significant keyword in the list (337.97) would not only artificially inflate the role of the KW but also contribute little of value to the overall analysis if included in the study.

5.2 KW classification by text function

How these terms are used stylistically is better indicated through Mahlberg and McIntyre's (2011) KW framework. As outlined in 2.3.3, this classification divides literary KWs by functionality, with fictional word signals building or otherwise contributing to representation of the text's fictional world and the thematic signals possessing greater thematic significance than their core denotational meaning (Mahlberg & McIntyre 2011: 209). Functional categorisation of these keywords thus enables visualisation of the representational methods through which *PP* approaches consumption on a lexical basis.

Figure 5.2 below illustrates the key fictional world and thematic signals involved in depicting consumption in *PP*. Through this framework the text's KWs are predominatly shown to be fictional world signals, which have been sub-divided into categories relating to 'characters', 'settings and props', 'temporal events', and 'actions'. By contrast to the collective fictional world elements, the thematic signals represent a smaller number of KWs and for this reason have not been classified further. The 'uncategorised' section at the bottom of Figure 5.2 contains the KW *brewer,* which does not presently fit into an existing category but would probably do so with further classification of the poem's KWs and the subsequent addition of new functional groupings; both Mahlberg and McIntyre (2011: 211) and Mastropierro

(2017: 96-7) observe that the presence of non-categorised KWs is not unusual, particularly when looking at a selection rather than all generated KWs for a given text or corpus. As with the initial classification of the poem's key themes in 4.2, selection as fictional world elements or thematic signals was based on analysis of the KW's concordance lines, with the function represented in the majority of its uses determining its categorisation in Figure 5.2.

Figure 5.2: Fictional world and thematic KWs in *PPCorp*

Category	Example key words
Fictional world	
Characters:	
Names	*hunger*, glutton**
Body descriptors	*lene*
Collective referents	*beggars, bakers, brewers*
Settings and props:	
Consumables	*food, drink*, ale, mete, bacon, bred,*
Temporal events:	*friday, lent, fridays*
Actions:	*drink*, beg, eat, begs, fasting*, begging*, starving*, y- dronke*
Thematic signals	
	hunger, glutton*, fruit, flesh moderation, temperancie*
Uncategorised	
	brewer

*denotes words split between different functional categories

The key lexis contained in the 'characters' category develop the fictional world of *PP* by identifying the individuals – and some of the bodily aspects of said persons –

inhabiting this text world. To this end this category includes *hunger* and *glutton*, which constitute the names of the allegorical characters Hunger and Glutton. The 'body descriptors' sub-category contains the the adjectival (and, in some instances, verbal) form of *lene* 'lean' characterising the potential bodily repercussions of non-consumption. The final sub-category 'collective referents' is more heavily populated with the plural common nouns *brewers, bakers*, and *beggars,* each which identify a collective group of persons involved in processes of consumption. *Beggars* is a particularly interesting term in this sub-category; unlike the traditional approach in *PP* of using collective referents to characterise groups of persons by profession (as seen through KWs such as *friars, doctors, brewers, priests*) beggars are generally characterised as individuals defined by the absence of profession, barring members of select religious orders.

Both the 'settings and props' and 'temporal events' categories are entirely comprised of nouns. Common nouns are used in the 'consumables' category to refer to edible products either with specificity (*ale, mete* 'meat', *bacon, bred*) or more generally (*food, drink*). By contrast, proper nouns dominate the events category as denoting days of the week (*friday, fridays*) or holidays (*lent*) in which consumption-based processes are altered for religiously-mandated reasons, generally via abstention from certain consumables.

The 'actions' category further develops *PP*'s fictional world through KWs alluding to the motive processes enacted by the poem's characters. These processes are typically represented by verbs describing the actions of consumption (*eat, drink, y-dronke*), with *drink* doubling as a potential nominal hypernym to liquid consumables like *ale* and *mede*. Additional verbs such as *beg(s)* suggest a textual focus on processes of non-consumption or consumption-based challenges, which are further supported by

terms such as *begging*, *starving,* and *fasting*; these KWs predominantly appear as particples but in some instances function as gerunds when describing these processes as more generalised states of existence under certain conditions (religious observances) and within particular demographics (e.g., the financially destitute).

By contrast, the KWs identified as thematic signals contain KWs representative of concepts having the potential to convey greater thematic significance beyond the creation of the text's fictional world in discussing consumption. This category contains the two concrete common nouns *fruit* and *flesh*, which convey different literal and figurative meanings in matters relating to consumption; their denotational meanings relate to dietary consumption, but the figurative meanings associated with *fruit* extend ideations of consumption to knowledge while *flesh* adds carnal overtones to consumption in select textual instances. This thematic signals category also features four abstract nouns which, like the gerunds in the 'actions' category, are reflective of broader existential states including *hunger, glutton*, *moderation,* and *temperancie* 'temperance'. These words have been identified as being thematic because they collectively signal the broader significance associated with consumption – beyond the physical processes of eating and drinking – on an ideological basis.

The stylistic potential exhibited in these functional KW groupings is particularly observable in the KWs split across these fictional world and thematic categories, denoted with an accompanying asterisk in Figure 5.2. Terms such as *drink, fasting, begging,* and *starving*, like with the split-words signposted in the semantic classification back in 4.2, lie on the intersecting boundaries between different categories, with their classification dependent on their contextual use. For example, *drink* is representative of two types of fictional world categories either as a generalised nominal consumable product (a drink), or the verbal process of consuming liquids (to

drink), but still remains a world-building element in the text. Alternatively, the functionality of terms such as *fasting, begging,* and *starving* as fictional world or thematic elements are fundamentally dependent on the contexts of their use as either verbal processes (someone is fasting as an action) or in more general descriptions as gerunds (fasting as a state of existence).

This focuses on three of these split-category KWs: *hunger* and *glutton* as examples of terms possessing dual fictional world and thematic functions (5.3), and *drink* as an example of textual multi-functionalism within a world building context (5.4).

5.3 Dual functionality via property allegory: *hunger* and *glutton*

The keywords *hunger* and *glutton* are stylistically unusual because their textual multifunctionality lies in their status as the allegorical characters Hunger and Glutton in *PP*. These KWs operate as examples of property allegory, wherein mental properties or ideological concepts (or, in the case of *hunger* and *glutton*, existential states) are used in the construction of a personified allegorical figure (Crisp 2001: 13-4). Property-based personification allegory presents a fictional character 'whose salient characteristic is that property' (Turner 1987: 21-2), with its evaluative potential based on being correlative to the given property; this means that virtually all characters are endowed with characteristics aligning with that of the property or ideological concept they represent. Glutton, for example, would not be presented as being satisfied or of diminuative stature but rather consistently characterised by excess and physical corporality.

This type of personification constitutes one of two rhetorical techniques in linguistic allegory. Personification involves the grammatical encoding of a noun or other part of speech's transformation into a proper name (Griffiths 1985: 5), with prototypical examples using abstract nouns to embody ideological concepts or properties. Personification is different from projection, a technique which involves little grammatical change indicating that content should be read allegorically, thus requiring more active readings of figurative significance based on relevant contextual knowledge brought to the text by the reader. By contrast to projection, personification is grammatically encoded in the language of the text, with its figurative meaning more visibly embedded in the lexico-grammatical patterns underlying its textual use. Despite differences in their linguistic realisation, both types of allegory operate as extended types of conceptual metaphor unique in their lack of explicit reference to a target domain (Crisp 2001: 9). From this perspective, it is unsurprising to find these KWs as thematic signals in *PP* alongside other examples of metaphoric language relating to consumption (e.g *temperancie*).

While most metaphoric language can be uncontentiously classified as thematic signals, personification allegory is not so clearly defined on a functional basis. It possesses a world-building function in its use as a proper noun identifying a fictional character by name (i.e., Glutton) but also operates thematically in the perpetuation of the broader ideological concepts imbued in the properties being personified (for example, themes of excess and gluttony within a literary text oriented around moral approaches to daily life). In other words, personification allegory is pragmatically understood to be thematic but possesses a distinct fictional word text function in its grammatical encoding not evident in other types of metaphoric language. This does not necessarily indicate underlying problems with Mahlberg and McIntyre's (2011)

framework itself but rather highlights a functional complexity and atypicality characteristic of allegory that warrants further investigation.

To explore how these instances of personification allegory stylistically contribute to representing consumption I first examine how these fictional world and thematic signals are linguistically encoded in *PP* through patterns in their surrounding collocates (5.3.1). I then, based on these collocational patterns, use Hoey's (2005) theory of lexical priming to offer insight into the ways in which the immediate context of an allegorical term can be used to prime readers to interpret the construct as being more literal (i.e., the name of a specific fictional character) or more figurative as an ideological property along a cline of readerly interpretation (5.3.2), including degrees of said interpretation (5.3.3). I finally complete my mini-investigation of *hunger* and *glutton* by comparing the textual patterns used to represent these chararcters in *PP* to the ones evidenced in the depiction of allegorical characters in the reference corpus texts to isolate any that are unique to *PP* (i.e. stylistic) as estranged from broader register-based allegorical conventions.

5.3.1 Collocations

In this section I first identify the main lexical and grammatical collocates for *hunger* and *glutton* in *PPCorp*. These collocates were generated through the concordance tool in *WordSmith* using mutual information (MI) with a L3-R3 window of proximity. Given that Hunger and Glutton are comparatively minor characters in *PP* – and the remnants of spelling variation are present in the collocational data – the minimum frequency of occurrence was lowered from the default 3 to 1 in order to capture all possible text patterns taking place around these KWs. The aim in generating these collocates is to

show these KWs to be part of larger textual patterns in the poem, which are then linked back to the different text functions (fictional world and thematic signals) in the next section (5.3.2). While discussion here focuses specifically on *hunger* and *glutton*, many of the textual findings raised in 5.3.1-5.3.4 are more broadly observable in the collocating patterns for other examples of personification allegory in the text (e.g. *conscience, mede, reason*).

As evidenced below in Tables 5.3 and 5.4, *hunger* and *glutton*'s main lexical collocates predominantly feature verbs and nouns. *Hunger* collocates most frequently with the verb *said*, but also with other verbs including *herkene* 'heed', *know, pray, suffer, save, take, a-wrekeme* 'avenge', *bade, sleep, heard, fed, grow*, and various forms of *to be* (*was, are, will, is*). This evidence of verbal collocation is also visible with *glutton* in *begins, go, goes*, and *y-globbed* 'gulped', with the high selection of phonetic parallelism present in these collocates for both *hunger* and *glutton* reflective of *PP*'s status as text written in intra-linear alliterative verse. By contrast, the nominal collocations for these KWs are a bit more varied, with *hunger* appearing in proximity to nouns including *sire, piers, haste, charity, people, mercy, master, man, work, hate* 'heat', *sickness, evidences, duntus* 'hits', *ale, hot, honger* 'hunger', *knight, folk, and god*; *glutton* similarly collocates with the nouns *sire, ale*, and *god*, and additionally appears alongside *morrow, galoun* 'gallon', *theft,* and *face*. Tables 5.3 and 5.4 additionally list a handful of adjectival collocates for each KW including *much, pure, good, blind, high, brokelelegged* 'broken legged', and *dentiesliche* 'deliciously' for *hunger* and *great, glad, guilty,* and *greedy* for *glutton*.

Tables 5.3 and 5.4: The 122 and 41 collocates for *hunger* and *glutton* respectively in *PPCorp,* organised by textual position relative to the node word and listed in order of collocational affinity by raw frequency.

LEFT lexical		RIGHT lexical
said (6), *sire* (4), *fear* (2), *charity* (2), *pure* (2), *god* (2), *herkene* (1), *piers* (1), *people* (1), *ybihote* (1), *please* (1), *suffer* (1), *evidences* (1), *mercy* (1), *folk* (2), *know* (1), *pray* (1), *fed* (1), *man* (1), *duntus* 'hits' (1), *ale* (1), *bade* (1), *behote* 'promise' (1), *hongur* (1), *knight* (1), *know* (1), *hot* (1), houped 'called out' (1), *garte* 'began' (1), *glad* (1), *god* (1), *folk* (1), *good* (1), *grow* (1)	*hunger*	*said* (2), *piers* (1), *haste* (2), *herkene* (1), *master* (1), *wend* 'walk' (1), *proud* (1), *know* (1), *work* (1), *save* (1), *take* (1), *sekennesse* 'sickness'(1), *dentiesliche* 'deliciously'(1), *sleep* (1), *blind* (1), *brokelegged* (1), *ate* (1), *hold* (1), *hate* (1), *heard* (1), *a-wrekeme* 'avenge' (1), *hold* (1), *high* (1), *hastes* (1)
LEFT grammatical		RIGHT grammatical
and (4), *the* (3), *for* (4), *that* (1), *in* (1), *of* (3), *not* (2), *are* (2), *y* 'I' (2), *hem* (1), *you* (1), *what* (1), *me* (1), *on* (1), *until* (1), *will* (1), *with* (1), *whether* (1), *such* (1), *dispit* 'despite' (1), *even* (1), *after* (1), *ac* 'but' (1), *also* (1), *be* (1), *any* (1), *his* (1), *is* (1), *from* (1)	*hunger*	*and* (9), *the* (3), *for* (2), *that* (5), *in* (4), *of* (1), *not* (1), *was* (3), *or* (3), *are* (1), *him* (3), *they* (2), *hem* (1), *you* (1), *here* (2), *what* (1), *tho* (2), *nor* (1), never (1), *this* (1), *then* (1), *shall* (1), *can* (1), *else* (1), *al* 'all' (1), *at* (1), *hitherwards* (1), *hit* (1), *if* (1), *hence* (1), *had* (1)

LEFT lexical		RIGHT lexical
great (1), *ale* (1), *morrow* (1), *sang* (1), *good* (1), *theft* (1), *sire* (1), *goes, gossip* (1), *god* (1), *begins* (1), *greedy* (1)	*glutton*	*great* (2), *ale* (2), *galoun* (2), *good* (1), *oaths* (1), *y-globbed* 'gulped' (1), *test* (1), *guilty* (1), *go* (1), *two* (1), *glad* (1), *face* (1)
LEFT grammatical		RIGHT grammatical
and (3), *a* (1), *you* (1), *y* 'I' (1), *and* (1), *the* (1), *now* (1), *to* (1), *then* (1), *until* (1), *also* (1), *even* (1), *hit* 'it' (1)	*glutton*	*with* (3), *a* (2), *to* (1), *have* (1), *y* 'I' (1), *and* (1), *you* (1), *would* (1), *me* (1), *in* (1), *for* (1), *had* (1)

In addition to these lexical collocates, *hunger* and *glutton* are shown to collocate with a range of closed-class grammatical words. These collocates operate less to describe or otherwise characterise the KWs but are rather reflective of the types of grammatical and syntactic constructs within which these KWs are situated in the text. To this end, *hunger* recurringly finds itself in textual proximity with conjunctions (*and, or, nor, until, if, whether, ac* 'but', *then*), personal pronouns (*him, y* 'I', *they, hem, you, me, hit* 'it', *here* 'their'), prepositions (*for, in, of, at, from, after, through, despite*), demonstratives (*this, that*), modals (*shall, can, could*), possessives (*his*), and some adverbs (*now, never, also hitherwards, hence*). Most of the grammatical collocates for *glutton* are repeats of those listed for *hunger*, with the exception of some definite and indefinite articles (*the, a*) and the modal *would*.

The larger number of collocates generated for *hunger* compared to *glutton* is linked to the number of raw occurrences of each KW in *PP*. *Hunger* appears a total of 33 times in the C-text of the poem (Wittig 2001: 305-6), with the current concordances generated for the KW identifying 32 of those instances, as set out below in Figure 5.5.

Glutton, by contrast, has only 10 total occurrences in *PP* (2001: 246), with the concordance illustrated in Figure 5.6 identifying 8 of these 10 textual uses.[57]

These grammatical and lexical collocates contribute to different underlying collocating patterns in the text. The main verbs, for example, primarily situate the allegorical construct (in this case, *hunger* and *glutton*) as the grammatical subject alongside verbs denoting physical movement, as in *hunger ate* (14 in Figure 5.5) and *goes glutton* (3 in Figure 5.6). This grammatical colligation with KWs as subjects also occurs through cognitive verbs conveying sensory actions like the hearing attributed to hunger in *hunger that heard him* (2), or else in the presence of verbs signposting dialogic speech most overtly visible in Figure 5.5 within the repeated co-occurrence of *said hunger* (20-24). Given the freer syntactic structure of Middle English these verbs – in addition to the L1 collocation evidenced in *said hunger* – are also shown to collocate in the R2-R3 positions to the allegorical referent in contructs such as *now begins glutton for to go* (1) and *for hunger hitherwards hastes* (8). A smaller proportion of these verbal colligations feature these allegorical constructs as grammatical objects, as exemplified in *fed hunger* (7).

[57] In both instances the disparity between total occurrences and those generated in the KW concordances was based on the presence of spelling variation for these words not picked up during the annotation process.

Figure 5.5: Concordance of 32 out of the 33 total occurrences of *hunger* in *PPCorp* in *WordSmith Tools*, ordered by L1 position

N	Concordance
1	nge Al for dread of here death such duntus ?af hunger Blind and Brokelegged he botened a thousand An
2	oghman y shall apayre ?ow all and houped after hunger that heard him at the first y pray the said Pe
3	n al my lifetime For y Vow to very god for any hunger or first shall never fish in the friday defyen
4	that makes ?ow to be sick Ac ate not y hot Are hunger the take And send the of his sauce ?sto sauery
5	o had Piers pity upon all pore people And bade hunger in haste high him out of contrary Home to his
6	ddares what best be to done For y know well be hunger went work they will full ill Meschef hit makes
7	am to chepyng And thennewere folk glad and fed hunger dentiesliche And then gloton with good ale gar
8	c y warn you workmen strive whiles you mow For hunger hitherwards hastes him fast He shall awaken th
9	cultivate Hit am y that find all folk and from hunger save Through the help of him that me hyder sen
10	ntiesliche And then gloton with good ale garte hunger sleep And tho would wastrel not work but wande
11	apese aloof prayed him blyue Have mercy on hem hunger said Piers and let me ?eve hem seeds And that
12	go abribeth and abeggeth and no man beten his hunger Piger propter frigus noluit arrare mendicabitu
13	ell is that never wo soffrede Nor what is hoet hunger that had never sin Ho could kindly whit colour
14	fered piers this present to please with hongur Hunger ate al in haste and asked after more Pore folk
15	greden after food and himself also suffer much hunger and wo in winter times And waking on nights to
16	ekeme of this wastrels for the knight will not Hunger in haste tho took wastour by the mouth And wro
17	spitteden and spradden fertilise in dispit of hunger They curuen here copes and courtepies hem made
18	friars of all the five ordres all for fear of hunger For that was bake for bayard was both for many
19	urs through thy might ac y pray the said Piers hunger are you wend Of beggars and biddares what best
20	me and also well ytauhte me y behote the said hunger that hence nor will y wend Are y have y-dyned
21	ia manducabis & cetera This are evidences said hunger for hem that will not swynke That here food an
22	y said Piers the ploughman ?e ybihote the said hunger or else the bible lies Go to our beginning tho
23	rink teach me now sire hunger Now herkene said hunger and hold hit for awysdom Bold beggars and big

24	work not so our womb greuethvs y wot well said hunger what sekenesse ?ow ails ?e have manged over mu
25	bour For here food and drink teach me now sire hunger Now herkene said hunger and hold hit for awysd
26	st y pray the said Perus tho pure charity sire hunger A-wrekeme of this wastrels for the knight will
27	ed with-sitte that Piers said for fear of sire hunger And Pereswas proud thereof and put hem all to
28	Ac ?ut y pray ?ow said Piers pure charity sire hunger if ?e can or know any kind things of physic Fo
29	pton on withfuales fro morning until even That hunger was not hardy on hem for to look For a put ful
30	ead nor devil die as god likes Whether through hunger or hate at his will be hit For if you live aft
31	such laws to learn labourers to grow Ac whiles hunger was here master there would non chide Nor dest
32	after crommes And ?ut had he esehem not for y hunger could him And affirm y say him sit as he a sir

Figure 5.6: Concordance of 8 out of the 10 total occurrences of *glutton* in *PPCorp* in *WordSmith Tools,* ordered by L1 position

N	Concordance
1	live with and to let you from theft Now begins glutton for to go to confession And carries him to ky
2	vphalderes an heep herly by the morrow G.even glutton with glad face good ale to hansull Clement th
3	lsedes for fasting days y bought hit Then goes glutton in and great oaths after Sesse the souhteres
4	rive and synege no more Y have good ale gossip Glutton would you test Have you said he any hot spice
5	d y and decretistre of Canon And also a greedy glutton with two great cheeks Hath no pity on vs pore
6	most easily should arise after And great sire glutton with a galoun of ale There was leyhing and lo
7	il euenson evensong and sang vm-by-while Until glutton had y-globbed a galoun and a gylle His goats
8	t a-loft art and all lives shaped To the god y glutton guilty y me ?elde Of that y have trespased wi

One of the emerging text patterns utilising both lexical and grammatical collocates for these KWs is binomials. The coordination of two words or phrases through syntactic elements such as 'and' or 'or' (Bhatia 1993: 108), binomials simultaneously involve collocation and colligation at the level of text based on the grammatical and semantic similarity shared by the two entities. These binomial patterns use the collocated conjunctions in either the L1 or R1 position to coordinate the allegorical referent with other common and abstract nouns. This construct does not appear in the 8 uses of *glutton* in Figure 5.6 but occurs twice for hunger in *hunger or first* 'hunger or thirst' and *hunger or hate* 'hunger or heat' (3 and 30 in Figure 5.5, respectively). While binomials are not shown to be highly common in the current dataset, examination of the concordances of the poem's other allegorical characters indicates that these patterns are recurringly utilised alongside these referents in the text, with higher overall numbers of binomials evidenced for personifications embodying abstract concepts (e.g. *reason, patience, mercy*) than states of existence such as *hunger* and *glutton*.

5.3.2 Lexical priming

How these collocates – and the text patterns within which this collocation takes place – assist in providing insight into the dual text functionality for words like *hunger* and *glutton* is informed by theories of lexical priming. Lexical priming (Hoey 2005: 13) postulates that individuals are experientially 'primed' to expect words to exhibit a preference to be used alongside other words (collocation), as well as within certain grammatical constructs (colligation). This priming on the linguistic level also contributes to the priming of different types of semantic and pragmatic meaning

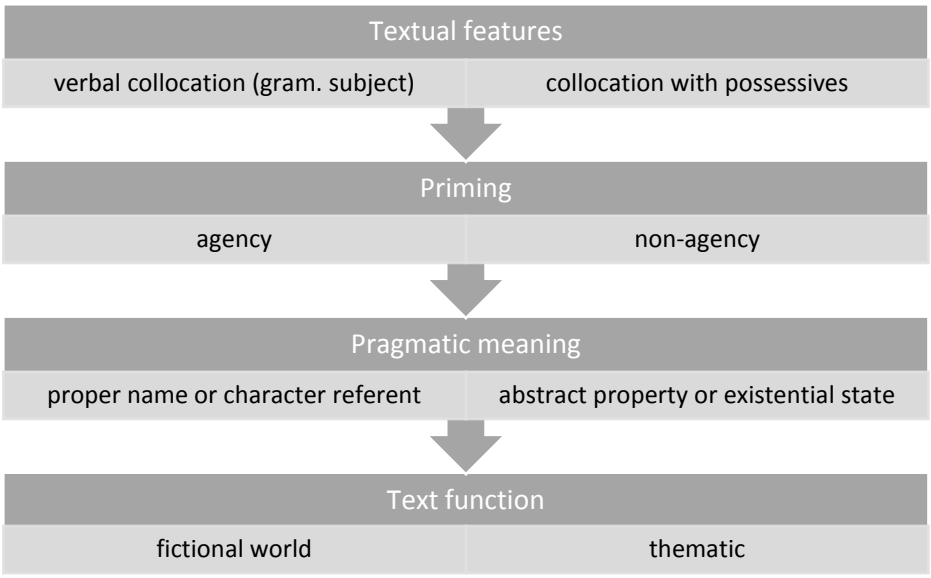
generated in the use of metaphoric language like personification allegory, with different collocating patterns guiding readerly audiences more towards a fictional world (i.e., interpretation of referent as character) or thematic functionality (referent as abstract concept).

A cumulative survey of the collocational patterns found for *hunger* and *glutton*, in addition to more general observation of the patterns exhibited in the concordances for the other key allegorical characters in *PP*, suggest that interpreting personification allegory is determinant on generated perceptions of agency. This agency is textually encoded in collocational and colligational patterns signposting personhood, with personified characters attributed with agency through the performance of their own independent actions, thoughts, and even dialogue (Yamamoto 1999: 10; see also Yamamoto 2006). In this sense, the verbal patterns associated with *hunger* and *glutton* act as 'carriers of personifying metaphor' (Leech &Short 1981: 198-9) which prime the application of an additional level of allegorical personhood to the given properties (in these cases, states of existence). This additional agency produces inferential meaning beyond what is typically expected of the abstract concepts of hunger and gluttony that facilitates their pragmatic interpretation as allegorical characters. To this end, agency operates as a generated effect which primes the textual functionality of *hunger* and *glutton* as world-building elements in the text; this process is more clearly illustrated below in Figure 5.7.

As such, it is the *absence* of primed textual agency that results in the production of a less distinct fictional world text function. The agency embedded in *hunger* and *glutton's* verbal colligation is not evident in its use alongside possessive pronouns, as exemplified in the phrase *his hunger* (12), which signals that the additional pragmatic layer of personhood is not meant to be added to the term's core

denotational meaning. This absence of layered inferential meaning instead primes a reading of the concept as the state of being hungry estranged from the allegorical significance embedded in more agency-based textual features. While possible for this POSSESSIVE + *hunger* patterning to indicate ownership of or a close relationship with another (in this case, allegorically constructed) person, this is not the case in (12) nor in virtually any of the examples wherein possessives co-occur with property allegory in *PP*. As a result, collocation with possessives thus primes readings of *hunger* as something to be possessed (or not) by people in general; to interpret this use of *hunger* as a referent to a character rather than a thematic signal would be fundamentally at odds with its underlying lexical priming.

Figure 5.7: Hierarchical relationship illustrating the way in which textual patterns – and the lexical priming resulting from said features – produces text function in property allegory, moving towards increasing levels of abstraction



Despite the neatness of the analytical levels represented in Figure 5.7, this binary divide between fictional world and thematic priming is an artificial construct from a cognitive perspective. In addition to problems associated with reading instances

of allegory too literally at the expense of its thematic significance (i.e, reading *PP* as being about a succession of dreams containing some very oddly-named characters) it is also reasonable to assume that textual patterns conveying non-agency are likely coloured by the background knowledge that the term might refer to the allegorical character despite the lack of overt markers promoting a fictional world function. This is particularly applicable to *PP* as a text operating within a larger allegorical literary tradition. In this sense the representation in Figure 5.7 is not meant to indicate that a dual understanding of a textual instance of personification allegory is impossible but rather that there are distinctive levels of textual and inferential processing that contribute to the interpretation of textual function in personification allegory.

5.3.3 Degrees of lexical priming

5.3.3.1 Animacy

How readers are primed to interpret the *degrees* of allegorical function is fundamentally linked to animacy. A cognitive distinction between the animate and inanimate operating along a hierarchical gradient (Yamamoto 1999: 149), animacy facilitates a generalised distinction between human entities and non-human concepts in allegorical constructs. It is the readerly perception of this animacy that underlies the search to explain the closeness experienced between reader and context in both Langacker's (1991: 305-29) empathy scale and Stockwell's (2009: 25) scale of attentional attractors in literary resonance. Both frameworks implicitly attribute the greatest amount of animacy to individual persons (i.e., an identified speaker or listener) identified in a text, followed by groups of people or indeterminate persons

and gradually leading to animals, objects, and abstract entities possessing the weakest links to the reader, as well as the smallest amount of recognisable animation. In this sense animacy is the product of conveyed agency and definiteness combined along a cline of anthropomorphic specificity, with it being natural to cognitively 'ascribe a stronger sense of animacy to an entity who/which is highlighted or activated as an individual in our minds than to one which is a part of an indeterminate mass' (Yamamoto 1999: 28, citing Comrie, 1989: 189). Studies such as Foley and Van Valin (1985: 288, cited in Yamamoto 1999: 27) supplement these cognitive understandings of animacy by examining the methods by which degrees of animacy are linguistically encoded.

The juxtaposition of these cognitive and textual approaches to animacy in Figure 5.8 presents a model for exploring personification allegory as the priming of varying levels of expressed animacy. From this perspective the different levels of animacy attributed to KWs such as *hunger* and *glutton* are clearly visible. Their forms as inanimate nouns representing abstract concepts inherently positions these terms at the lowest end of the animacy scale on the far right. This classification changes markedly with their change in use to proper nouns identifying specific human persons, which reorients the same word to the opposite end of the animacy scale.

Figure 5.8: Complementary theories reflecting animacy in language along a gradient of decreasing animacy from left to right, with [>] denoting sub-levels of animacy within each grouping

Cognitive animacy scale	***Langacker (1991)***	Speaker> hearer	Human	Animal	Physical object	Abstract Entity
	Stockwell (2009)	Specific persons	Ill-defined individuals >groups	Animals >plants	Machines> immovable objects	Landscape objects> abstractions
Interpretative animacy scale		Agency (movement>speech>sensory> sentience)			Absence of agency (possession)	
Linguistic animacy scale	***Foley & Van Valin (1985)***	1st & 2nd person>3rd person pronouns> proper nouns	Human common nouns	Other animate nouns	Inanimate nouns	

In addition to providing a method of identifying the levels of semantic animacy within *hunger* and *glutton* as decontextualised lexical units, Figure 5.8 also illustrates how the priming of each KW as fictional word or thematic elements is influenced by the animacy in their collocates. This is revealed in the interpretative scale, which connects the cognitive and linguistic scales in acknowledging the role played by collocational patterns on pragmatic inference. For example, the high animacy resulting from the recurring L1 collocation to the noun *sire* (25-28 in Figure 5.5 & 6 in Figure 5.6) demarcates these entities as specific persons, with the use of a title here signposting the use of *hunger*/*glutton* as proper nouns, and hence world-building elements as the names of recognisable characters.

Applying interpretative animacy beyond semantic encoding to the textual features and patterns surrounding these KWs reveals the comparative levels of thematic and fictional world priming operating within individualised examples of personification allegory. As discussed in 5.3.2, patterns like verbal colligation layer a

fictional world function over an underlying thematic function. The strength of these world-building versus thematic functions relative to one another is based on the agency-based animacy contextually encoded in the different types of collocates surrounding the use of these KWs.

To better differentiate between these types of conferred agency, Table 5.9 below takes the interpretative animacy scale set out in Figure 5.8 and aligns it with examples drawn from the concordance lines for *hunger*. This step reveals the collocate-based impact on allegorical interpretation as a cline wherein collocates ascribing physical movement characterise the entity with the highest amounts of personhood (and thus, the strongest fictional world signal), followed by collocates conveying speech, sensory information, and finally expressions of sentience. To this end, *ate* (i) operates alongside other motive verbs such as *go*, *goes* (1 and 3 in Figure 5.6), and *took* (16 in Figure 5.5) in conveying the most agency to *hunger* through physical motion relative to other highly animate verbs denoting articulation through asking (i) and saying (ii); these two processes collectively represent the most common methods of signposting personhood – and thus, a fictional world functionality – for *hunger* in *PP*.

Table 5.9: Levels of agency attributed to *hunger* via collocation, in descending order of encoded collocate-based animacy

	Collocate		Example (ModE)	Text function
Movement	*ate, asked*	(i)	Hunger ate it all and asked for more (14) (8:319)	Fictional world
Speech	*said*	(ii)	I promise you, said hunger, I won't go away/ before I have this day both dined and drunk… (20) (8:301-2)	Fictional world
Sensory	*heard*	(iii)	Now by Christ said Piers, I'll punish you all/ and called after hunger who heard right away… (2) (8:167-8)	Fictional world
Sentience	*know*	(iv)	I know well, said hunger, what sickness ails you… (24) (8:270)	Fictional world
Possession	*his*	(v)	The sluggard that grows no corn because of the cold/ In the summer for his sloth shall suffer want/ and go a-begging and begging and no man abates his hunger… (12) (8:244-6)	Thematic

The next tier of collocates (still conveying personhood, but less animatedly) contains verbs denoting cognitive processes (iii and iv in Table 5.9 above). These cognitive processes still convey agency to states such as *hunger* though reference to sensory experiences like hearing (iii) and the general sentience denoted in knowing (iv) but not to the same degree facilitated through colligation with motive verbs or overt speech acts. The final example (v) is interesting in that, while the pronoun *his* itself contains a high amount of animacy linguistically, this animacy does not carry over pragmatically when collocating with *hunger*; what does get conveyed is the possession of said state, which ultimately results in the KW retaining its non-allegorical thematic functionality.

5.3.3.2 Foregrounding and commonality

To this end, interpretation of *hunger* and *glutton* as allegorical referents is fundamentally based on the foregrounding of fictional world functionality over the thematic. This foregrounding effect (Leech 1968, Short 1981) is based on the underlying priming of these KWs, wherein one function is rendered more salient than the other even if interpretation remains grammatically possible for both functions. In most cases for *hunger* and *glutton* both text functions are generally accessible unless primed with sufficient strength to make a reading based on the other text function implausible (e.g, the thematic functions encoded in markers of possession). This functional duality in textual interpretation is fairly common, and is particularly visible in the description of Hunger's actions against the people reluctant to farm the land in passus 8:

(i) Hunger in haste tho hente wastour by þe mawe
And wronge hym so by þe wombe þat al watrede his yes… (16)

Hunger in haste then grabbed waster around the belly
And held him so tight that his eyes watered… (8:171-2)

As illustrated above, the agency occurring in *hunger…hente* primes an allegorical understanding of *hunger* as a person physically grabbing someone in a tight embrace. However, its underlying thematic functionality is still visible, making it possible to simultaneously envision *hunger* as an existential state plaguing someone with eye-watering stomach cramps and pains. In these instances, the core denotational meaning of *hunger* is still present and able to be interpreted by readers but it is far more difficult to *exclusively* understand as referring to the ideological concept; in other

words, an exclusively thematic-based figurative reading is still possible but vastly more cognitively challenging.

This foregrounding of text function is also visible in the binomial patterns underlying *hunger*. The grammatical and semantic similarity evidenced in these binomials signals functional foregrounding based on the concept coordinated with the text's allegorical referents. For instance, the coordination of *hunger* alongside common nouns such as *thirst* and *heat* (3 and 30 in Figure 5.5) are indicative of a thematic functionality; had either of these terms been used as a proper noun, the KW's functionality would have reversed accordingly in the new primed understanding of two abstract human-esque entities.

This foregrounding of one text function over another moreover illustrates how pragmatic meaning is extrapolated from textual occurrences characterised by ambiguity. Occurring when the underlying priming does not clearly prioritise one text function over the other, this pragmatic ambiguity makes the reader work harder to extrapolate meaning from potential allegorical constructs. This absence of foregrounding at the level of text is predominantly observable in *hunger* within prepositional phrases such as *through hunger* (30) and *fear of hunger* (17, 18). The preposition *of* in both instances denotes *hunger* as the source or primary cause of said fear, but with little additional guidance on whether to interpret these examples allegorically or literally. *Through* is similar in its potential reference to a process mitigated by (or conducted through) a physically-denoted person versus a cognitive capacity. Some of these ambiguous cases are assisted by contextual clues in the surrounding text, however this is not always the case.

Aside from the textual encoding of animacy, the interpretation of allegorical text function also relies on the relative commonality of these primings. Allegorical

extrapolation does not occur in vacuum conditions limited to the given text (in this case, *PP*) but additionally integrates broader consideration of the relative frequency by which a given word, compound word, or verbal phrase is used allegorically versus non-allegorically in general language use. This frequency is experiential as well as textual, with 'the rarer sense gets primed in such a way as to ensure that their primings do not overlap with those of the most common sense' (Hoey, 2005: 104). The variant, or non-prototypical, nature of personification (Stockwell, 2002: 60) in words such as *hunger* and *glutton* outside a limited corpus of historical literary and religious English texts results in their non-allegorical interpretation on a general basis unless otherwise signposted textually. This commonality is also dependent on modern versus medieval readings of allegory, with a medieval readership operating within the text's originating discourse world (Gavins, 2007: 9-10) likely possessing a stronger experiential priming (not to mention a substantively different prototypical conception system) for allegorical interpretation than the modern audiences for whom personification allegory is no longer an expected feature of modern poetic language use.

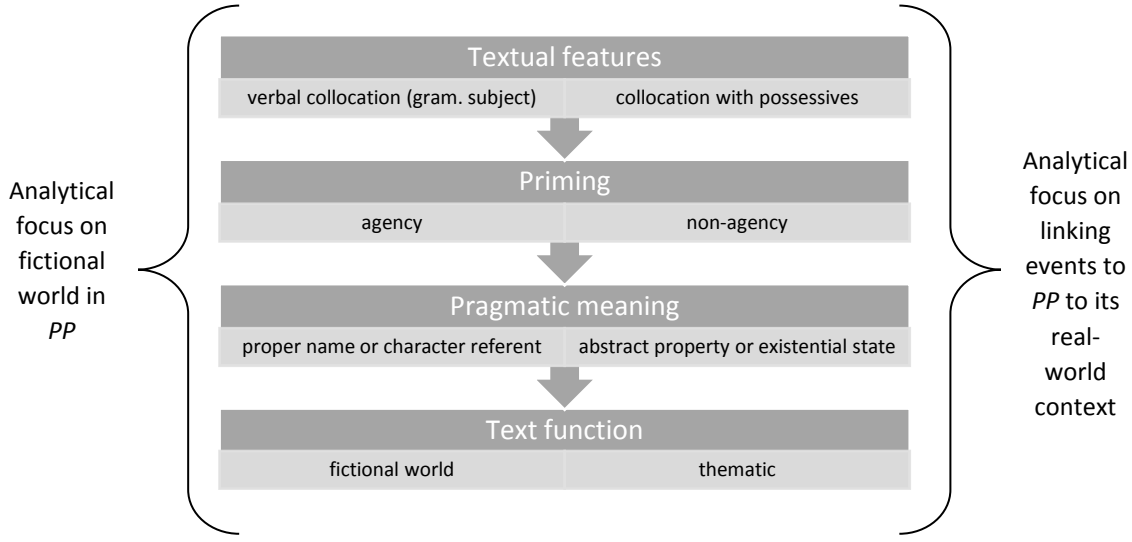
5.3.4 Hunger and glutton in the literary criticism

The significance of the CS analysis conducted in 5.3.3.1-5.3.3.2 is that it corroborates and further builds on the findings found in the existing scholarship surrounding *PP*. My corpus data confirms observations of dual literal and figurative meanings that can be extrapolated from consumption-based allegorical figures (e.g. Mann 1979), as revealed in the comparative priming of thematic and fictional world functions revealed in the KW concordances. My subsequent analysis of this corpus data moreover offers ways by which to explain why multiple interpretations of key concepts in *PP* such as *hunger*

and *gluttony* have persisted in the critical scholarship, and how this interpretative divide is an organic result of studying texts containing high amounts of metaphoric language use.

Within the context of the current study, the conflicting interpretations offered for personification allegory like *hunger* can be reframed as the divides between which text functions are being foregrounded in the critical analysis of the text by each scholar. For example, it was observed back in 2.1.6.2 that Langlandian scholars offer different readings when larger ideological challenges are physically manifested in the text of the poem, with Savage (1993) characterising consumption as being a struggle between internal or external forces and Kim (2004) orienting this struggle as one area in a larger network of political conflict between Christianity and feudalism. These differences in stance take on additional meaning when analytical direction is superimposed on the language patterns driving this distinction between literal and figurative interpretations (originally shown in Figure 5.7), as reproduced below as Figure 5.10:

Figure 5.10: Reproduction of Figure 5.7 linking interpretative trends in the literary criticism to the directional focus corresponding to each analytical aim.



In light of the distinctions outlined above in Figure 5.10, Savage's (1993) reading is based in the fictional world of *PP* as a struggle between the internal and external forces within the fictional characters themselves. By contrast Kim's (2004) reading of political conflict, despite having thematic resonance with the events occurring in the text, more fundamentally operates as an interpretation that engages with the ongoing real-world economic and political strife taking place in the late 14th century context of *PP*'s textual composition. In this sense the critical scholarship in *PP* studies exemplifies Werth's (1994: 102) description of literary ambiguity as something which 'splits its audience into two groups: those who see one meaning and those who see another', without either of these groupings having to be characterised as erroneous depite the binary nature of their interpretative stances.

Some critical interpretations of the poem notably 'toggle' between fictional and thematic interpretations (figurative and literal meanings, respectively) of the poem's KWs. Gavins (2007: 152) originally uses 'toggle' in TWT to illustrate how participants shift between source and target domains to ascertain meaning in metaphoric constructs, with participants 'able to toggle between the worlds if necessary'; however, given the absence of a target domain, toggling in this sense is not particularly useful to interpreting meaning attributed to allegory unless it be in establishing personhood from thematic significance. Instead, my use of the term here extends this same application to the action wherein participants fluctuate between the primed meanings of a word (rather than the target and source domains in a specific metaphor) to achieve the same end of ascribing meaning to the literary metaphor present in individual textual examples of personification allegory.

This type of 'toggled' interpretation can be seen in the readings involved in the debate surrounding the character Hunger, which pit the literal significance of physical

hunger as social critique against the more figurative understanding of hunger as a process of spiritual exegesis (Benson 2004: 138, Gasse, 2016: 67-8). In this case, unlike the Savage (1993) versus Kim (2004) opposition, these interpretations actively toggle between the two interpretative stances in the search for information. Benson (2004) and Gasse (2016), for example, do this by taking 14th century conventions of social critique and biblical exegesis and applying them to the actions attributed to characters within the poem's fictional world. This ultimately corroborates Mann's (1979) observations that multiple permissible interpretations are possible in Langland's representation of allegorical consumption, but further build on her assessments by revealing some of the underlying technical, textual, and cognitive mechanisms through which these various allegorical readings arise.

The following section (5.3.4) concludes the present discussion of *hunger* and *glutton* by looking at the use of these KWs in the study's reference corpus to determine if the identified patterns of use found in *PP* are stylistic, or correspond to more broadly recognisable patterns indicative of the text's poetic register.

5.3.5 PPCorp vs. MEPoetLitCorp

By contrast to the highly stylised uses of *hunger* and *glutton* in *PPCorp*, these words are overwhelmingly used as thematic signals in the reference corpus. The 6 occurrences of *glutton* in *MEPoetLitCorp* reveal recurring L1 collocation with the (in)definite articles *a* and *the*, which identifies a single person via a particularly notable characteristic attributed to said individual. In this respect *glutton* is not unlike the KW *gambler* in the Bond novel *Casino Royale*, which is classified as a thematic signal (Mahlberg and McIntyre 2011: 209). *Hunger* is also fundamentally more thematic in

MEPoetLitCorp, with 7 of the 39 total occurrences appearing in L1 collocation with the possessives *his* (5) and *my* (2). The predominant type of collocation with *hunger* is shown to occur 16 out of 39 times within prepositional constructs such as *with hunger* (3)*, for hunger* (10)*, of hunger* (2)*,* and *by hunger* (1) which, on closer examination, all refer to hunger as the existential state.

There are a few ocurrences in *MEPoetLitCorp* wherein *hunger* is shown to appear within collocations signalling greater amounts of agency. As seen with the use of *hunger* in *PPCorp*, verbal collocates convey agency to *hunger* in several instances within the reference corpus, as illustrated below in (ii-iv):

(ii) I am so hungry that I may not sleep,
Now would god that I might sleep ever
Than should not hunger in my womb **crepe**… (*MonT*)

(iii) Until I may have the great feast,
Which al my hunger might **arrest**. (*CA*)

(iv) Such thirst and hunger him **assails**
That nevere his appetit nor fails. (*CA*)

While perhaps not enough to convey recognisable personhood as separate human-esque entities like Hunger and Glutton in *PP*, the collocation of *crepe* 'creep', *arrest*, and *assails* act as 'carriers of personifying metaphor' (Leech & Short, 1981: 198-9) in creating the notion of independent movement via high amounts of animacy. This process is particularly observable in *Confessio Amantis* (*CA*), which also contains allegorical elements like *PP*.

However, despite this verbal colligation, none of the uses of *hunger* in the reference corpus retain a fictional world functionality. The world-building dimension attributed to *hunger* in *PP*, while gathering stonger amounts of agency via verbal colligation, is fundamentally based on the KW being used as a proper noun signifying

the allegorical name of a specific character, a process which is not evident within any of the texts in the reference corpus. Using verbal collocates to produce agency for *hunger* thus results in highly figurative but not allegorical meaning; in other words, all instances of personification – including those indicated via collocation or colligation – are metaphorical but they are not all allegorical (Stanley 2008: 21). To this end, the figurative conceptualisations of *hunger* evidenced in (ii-v) are not world-building elements but rather function as stylised thematic signals retaining atypical levels of personification-based significance.

This comparison of the use of *PPCorp* to *MEPoetLitCorp* indicates that the stylistic significance in the use of *hunger* and *glutton* have more to do with the relative frequency with which they are used as fictional world elements, rather than changes to the underlying language patterns conveying each function. *Hunger* is not personified in any of the reference corpus texts, nor is *glutton*.[58] However, despite this absence of identifiable personifications (which do not require a corpus analysis to identify), this corpus comparison did reveal certain language patterns revealing split-functions where fictional world interpretations *could* be interpreted, even if not contained in an 'allegorical' poem. This is a characteristic feature that I strongly suspect is characteristic of many of the other KWs generated for the poem, and would serve as an interesting analytical focus for future studies (see Chapter 7).

The final part of this chapter (5.4) continues the present focus on words of split functionality by examining the use of *drink* as a term possessing different types of world-building functions in *PP*.

[58] *Gluttony* does play a minor role in *CA* as one of the seven deadly sins, but does not appear in a concordance search for the KW *glutton*.

5.4 Dual functionality in fictional world development: *drink*

The third consumption-based KW being examined in this chapter is *drink*. As observed in 5.2 *drink* is used in *PP* as both a noun identifying a type of consumable, as well as a verb conveying the action of drinking. Unlike *hunger* and *glutton*, *drink* operates as a bridging word within different sub-categories of the same text function (fictional world element). This ssection sets out to examine the different text patterns underlying the two ways in which *drink* contributes to the development of the fictional world in *PP* as both a consumable and a performable action, and to explore its stylistic significance to the text. *Drink* is less overtly stylised compared to *hunger* and *glutton* – predominantly through lacking the additional figurative significance created by the latter being allegorical referents – but still constitutes a term of interest by virtue of being a 'split' word (see Figure 5.2).

This section broadly mirrors the approach taken with the other KWs in 5.3. 5.4.1 first examines the collocates for *drink* using *WordSmith*, which reveals the recurring text patterns underlying its use as an action versus a consumable product as well as high amounts of semantic and grammatical negation present in said collocations. 5.4.2 examines how this collocate-based negation reveals a strong evaluative approach to the concept of drink(ing) in *PP,* with reflective discussion on the link between evaluation and consumption more generally. 5.4.3 contains a more detailed investigation of this negative evaluation by using Martin and White's (2005) classification system for evaluative language. This approach contextualises how drink(ing) is framed in *PP* – particularly through valuations of propriety (5.4.4) – and also enables a more linguistically, ideationally, and culturally nuanced comparison to the ways in which *drink* is represented in the reference corpus texts (5.4.5). This comparison of *drink* in *PPCorp* and *MEPoetLitCorp* ultimately offers a method of

ascertaining whether the methods by which *drink* is evaluatively portrayed in the poem are stylistically unique or representative of larger discoursal practices as register features.

5.4.1 Collocations

The top collocates for *drink*, as generated from the annotated version of *PPCorp*, are reproduced below in Table 5.11. As with the approach taken with *hunger* and *glutton*, these collocates were generated using MI using a 3-word span on either side of the KW in at least one occurrence. These collocates are listed below in order of their measured affinity to *drink,* with raw frequencies denoted in parentheses.

Table 5.11: The 64 total lexical and grammatical collocates of *drink* in *PPCorp*, organised by textual position relative to the node word and listed in order of collocational affinity by raw frequency.

LEFT lexical		RIGHT lexical
men (2), *love* (2), *gluttinous* (2), *death* (1), *day* (1), *mete* (2), *find* (1), *lack* (1), *dread* (1), *live* (1), *noon* (1), *toke* (1), *turn* (1), *times* (1), *souls* (1), *sweet* (1), *breed* 'bread' (1), *delightful* (1), *delicatlyche* (1), *dregs* (1), *dine* (1), doctor (1), *derworth* (1), *ale* (1), *broke* (1), *lest* (1), lent (1), *lykerous* (1), *moder* (1), *feeble* (1), *Iosophat* (1), *good* (1)	*drink*	*men* (2), *death* (2), *day* (2), *wicked* (2), *body* (2), *deed* (2), *find* (1), *lack* (1), *dread* (1), *live* (1), *pain* (1), *reach* (1), *right* (1), *worth* (1), *women* (1), *wynninges* (1), *wrought* (1), *wine* (1), *buxumnesse* (1), *sleep* (1), *ripe* (1), *send* (1), *teach* (1), *taverns* (1), *deserven* (1), *deserve* (1), *died* (1), *diverse* (1), *delicately* (1), *diapseurans* (1), *dies* (1), draught (1), *drive* (1), *deep* (1), *beg* (1), *beggars* (1), *blame* (1), *life* (1), *lene* (1), *lote* (1), *fail* (1), *evening* (1), *druie* (1)

LEFT grammatical		RIGHT grammatical
and (21), for (5), ne (6), of (4), to (5), y (4), in (3), his (6), that (2), þe (3), ac (2), hit (1), the (1), a (1), thy (3), there (1), he (1), but (1), my (3), at (1)	*drink*	and (17), for (6), the (4), ne (3), of (5), to (3), y (4), in (5), his (1), that (4), ac (2), hit (3), no (4), a (2), there (2), he (2), but (2), as (3), at (2)

In looking at Table 5.11, the lexical and grammatical collocates each reveal different aspects of the KW's use in the text. In addition to collocates signposting the poem's alliterative metre (*day, delightful, delicatlyche, dregs, draught, diverse, deserve*(*n*), *dine*), the lexical collocates reflect the KW's dual nominal and verbal functions in the presence of terms referring to agents of drinking (*men, beggars*) as well as types of drink (*ale, wine*), which appear alongside spatio-temporal words denoting the settings in which consumption does or does not take place (*taverns, day, noon, lent, evening*). This functional distinction between *drink* and *drinking* is more clearly visible in the term's grammatical collocates, particularly with collocates appearing in the L1 position; patterns such as POSSESSIVE PRONOUN (*his, thy, my*) + *drynke* and ARTICLE (*the, a*) + *drink* signal its meaning as the beverage, while PERSONAL PRONOUN (*y* 'I', *he*) + *drink* and *to* + *drink* as a verbal infinitive signposting discussion of drinking as an action. The grammatical collocates calculated to have the greatest affinity to the keyword – *and, for, ne,* and *of* – offer little overt signification of *drink*'s functional meaning, with either possible in grammatical constructs such as binomials (*and, ne*) or in isolated phrasal elements (*of, for*).

Despite its unusually high collocational affinity to *drink* in the annotated dataset (28 times), the lexical collocate *food* is not included in Table 5.11, nor is it considered in the present investigation. It consistenly occurs in the annotated dataset in an L1 position resulting in the binomial *food and drink*; however, this construct does

not appear anywhere in the actual text of *PP*. The presence of this pattern in Table 5.11 is instead based on the original corpus annotation wherein the ME term *lyflode* was modernised to 'food and drink'. This textual intervention caused the 28 occurrences of the artificially separated components of *lyflode* to appear alongside the genuine uses of *drink* (originally *drynke*) in *WordSmith*'s concordancing tool as revealed below in Figure 5.12.

Figure 5.12: Screenshot of the XML tagging embedded in the annotated *PPCorp* illustrating the lexical origins of *food and drink*



To this end, my analysis of *drink* extends only to the collocates and concordances in which the appearance of *drink* is replicated in the original source-text. This means that out of the 64 total instances of the term identified by *WordSmith* only 36 are used to examine the text patterns underlying *drink*'s text functionality. These 36 occurrences are set out below in Figure 5.13, from which all concordances featuring *food and drink* (2-3, 5-14, 16-18, 20-28, 30, 33-35) have been omitted.

Figure 5.13: Concordance of 36 out of the 64 total occurrences of *drink* in *PPCorp* in *WordSmith,* ordered by L1 position

N	Concordance
1	y ate more fish then flesh there and feeble ale drink Ac other while when wine comes and when y dri
4	nd bad hem souke for sin save at his breast And drink both for bale brouke hit ho-so might O vos om
15	l the ventage valle in the vale of Iosophat And drink right ripe m.ust resureccio mortuorum And the
19	Fro Mihel-masse to Mihel-masse y find mete and drink y find pain for the pote and preen him i woul
29	es The first is food and vesture the second And drink that doth the good and drink not out of time
31	ure the second And drink that doth the good and drink not out of time Loot in his live through glut
32	ristene And fasting days to eat before noon and drink With spiserye speak idleness in vain speak an
36	that you know by countenance nor by speech And drink not over delicately no to deep neither That t
37	e Or ?e shall ate barely breed and of the broke drink but if he be blind or brokelegged or bolted w
38	now brouk hit yourself That art doctor of death drink that you made For y that am lord of life love
39	here was loot assailed For-thy dread delightful drink both day and nights Measure is medicine thoug
40	or sweet thing assuage my swelling Nor derworth drink drive hit fro my heart Nor neither shame nor
41	ard was both for many hungry drosenes and dregs drink for many beggars There was no led that lived
42	the not care for no corn nor for cloth nor for drink Nor dies dread nor devil die as god likes Whe
43	out of time Loot in his live through gluttonous drink Wicked wrought and wrathed god almighty In hi
44	pole poysen potten up to his lips And beden him drink his death to let and his days length And said
45	submissive at his bidding for his breed and his drink And buxumnesse and be are evermore at werre A
46	ankind In mete out of measure and many times in drink In women in weeds and in words both They ouer
47	l said that shrive saturdays for thy moder love Drink but with the duke and dine but ones Y guilty
48	easure That many a lede leases through lykerous drink As lote deed and No and herodes the daffe ?af
49	is physic this folk hath enchanted And doth men drink dwale that men dread no sin By Christ said Co
50	you made For y that am lord of life love is my drink And for that drink to-day y died as hit seeme
51	er Movable possessions nor vnmeble mete neither drink Ac then deed we all well and well bet ?ut to

ttle silver Nor doth men dine delitcatlyche nor drink wine oft A straw for the stuyues hit stoed no
ain here finds And moderation ?eueth here sweet drink and entertain here in all angrys Thus learned
t am lord of life love is my drink And for that drink to-day y died as hit seemed Ac y will drink o
 of communitу coppes all christian souls Ac thy drink worth death and deep hell thy bolle y fauht s
e Conscience by so it cache silver Then y do to drink a draught of good ale And so said sixty of th
a gives hem and great oaths together All day to drink at diverse taverns And there to iangele and t
th and loved well fare And no deed to do but to drink and to sleep In hele and in inwitt one me app
at his ease Rest him and roast and his rug turn drink druie and deep and draw him then to bed And w
and said Agite penitenciam And sethe he drow vs drink diapseuerans As long said he as life and body
 clean wheat Nor noon halfpenny ale in none way drink Both of the best and of the brownest that bre
hat drink to-day y died as hit seemed Ac y will drink of no dische nor of deep clergyse Both of com
drink Ac other while when wine comes and when y drink late at evening Y have a flux of a foul mouth
 sire sorfeet sittien at thy board And look you drink no day are you dine somewhat And think that .

In returning to the collocates generated for *drink*, one of the striking features revealed in Table 5.11 is the high amount of negation evidenced in these collocates. Several of these grammatical words are overtly negative, as seen with the articles *ne* (meaning 'no', 'not', or 'nor', depending on its use) and *no*. *Ne* appears nine times in various syntactic positions to the right and left of *drink*, with *no* collocating an additional four times; even when not directly co-occurring with the KW, the proximity of these grammatical collocates often ascribes negativity to aspects linked to drink(ing). This negation produced by their recurring collocation is particularly visible with the conjunction *but* and its more historical form *ac*; they convey less overtly recognisable negation semantically but rather achieve this grammatically through the negative semantic prosody generated by clausal positioning when syntactically preceding the KW. A typical example of this type occurs in (58), when the narrator awakens as a worker disinclined to labour in the fields:

(i) In an hot heruest whenne y hadde myn hele
And lymes to labory with and louede wel fare
And no dede to do **but to drynke** and to slepe (58)

In a hot harvest when I had my health
And strong arms to work with but loved the good life
And no deed to do ***but to drink*** *and to sleep.* (5:7-9)

In contrast to this negation via grammatical collocates, the lexical collocates appear to ascribe negativity to *drink* through the embedded orientational metaphors operating within these collocating terms. Terms such as *death* (38, 44, 55), *dies* (42), *died* (54), and *deep* (59) convey negation in the downward spatial orientation of their metaphorical use (Lakoff & Johnson 1980a: 14-7); *deep* offers a literal interpretation of this, with references to death being down in relation to life from a more existential

perspective. This poetic technique of surrounding a KW not encoded with its own overt evaluative meaning with metaphorically negative terms can also be seen in the collocation of *drink* with *dread* (39, 42), which conveys a negatively emotive metaphoric orientation. Lines containing either multiple instances of lexical negation, or combinations of grammatical and lexical methods, produce particularly colourful representations of the unfavourable things associated with *drink,* as reflected within God's speech to Lucifer in (38):

(ii) The bitternesse þat thow hast browe now brouk hit thy-sulue
That art doctour of **deth drynke** þat thow madest…
Ac y wol **drynke** of **no dische ne** of **deep** clergyse
Bote of comune coppes alle cristene soules
Ac thy **drynke** worth **deth** and **depe** helle thy bolle. (38)

The bitterness that you have brewed, now enjoy it yourself
You who are doctor of ***death, drink*** *what you've made…*
But *I will* ***drink of no dish*** *of deep learning*
But from the common cups of all Christian souls
But *your* ***drink becomes death, and deep hell*** *your bowl.* (20.414-5, 418-20)

This does not imply that there are no positive representations of *drink* in *PP*, rather that the keyword's collocational relationships reveal a pervasive negativity associated with drink(ing) in the text.

One pattern in particular emerges in the poetic use of this collocational negation relative to the text's alliteration. Out of the 36 total occurrences of *drink* in *PP* the word appears 21 times within the set alliterative pattern for each line, with the remaining occurrences textually positioned outside of the line's primary alliterative scheme. It is interesting to note that, from among these 21 textual instances wherein *drink* contributes to the line's alliterative pattern, 16 also contain some form of negation. This negation is manifested either from the keyword's grammatical or lexical

collocates (or, in instances like (ii), from both types), while only two examples (24, 35) of *drink* appearing in non-alliterative contexts feature any sort of intra-linear negation.

This pattern highlights a distinctive preference for the representation of drink(ing) in *PP* to be fundamentally linked to the combined use of alliteration and negation simultaneously, a recurring formula which serves to stylistically foreground this negative view ascribed to *drink* within the text. The unusual repetition of word-initial phonetic repetition and syntactic parallelism characteristic of alliteration renders these patterns – and the words within those patterns – more perceptually prominent, which serves to amplify this negative colouring of drink(ing). In other words, the observed semantic preference of *drink* to appear proximal to orientation-based conceptual metaphors amplifies the negative semantic prosody associated with drink(ing) in the poem, negativity further emphasised by the inclusion of these negated elements within the text's alliterative scheme. This approach could well function as one of the reasons behind the overwhelming prosodic negativity attributed to drinking in the poem, despite the term's own relative semantic neutrality in contextual isolation.

5.4.2 Evaluative language and consumption

While textual patterns conveying negative polarities are generally more perceptually prominent than positive ones (e.g. Channell 2000: 55), the negation characterising the use of *drink* exists as part of a larger correlation present in the original list of consumption-based KWs generated for *PPCorp* back in Figure 5.2. While less visible under decontextualised conditions with terms in the 'consumables' category like *fruit, bred,* and *mete*, many of these consumption-based KWs contain inherent evaluative

content where negation is 'inscribed' in the meaning of the term to the degree where its recognition is a given observation (Hunston 2011: 58, citing Martin & White 2005). This is particularly applicable to the terms found in the 'actions', 'body parts', and 'thematic signals' categories, especially within the KWs signposting existential states and descriptors. In the case of states such as *hunger* and *starving*, this embedded negation is based on the underlying correlation wherein a lack of recourse to necessary sustenance as a potentially life-threatening condition or situation is associated with being 'bad' as a type of orientational metaphor where HEALTH AND LIFE ARE UP and SICKNESS AND DEATH ARE DOWN (Lakoff & Johnson 1980a: 15). This evaluative understanding extends to consumptive states where sickness and death are perhaps not yet present but foreseeable if said state or condition continues unchanged, with the bodily characteriser *lene* also able to be interpreted with this type of fundamental negativity when associated with insufficient nourishment. *Glutton* is similarly negative but for reasons of consumptive excess rather than the privation observed in the other three KWs. *Moderation* and *temperancie* are particularly interesting terms from an evaluative perspective; in representing habitual restraint in consumptive appetites (dietary or otherwise) these terms are superficially positive as states of existence via the amount deemed appropriate for everyday consumptive practice, with the fluctuation from moderation to either side attributed with negativity through the perception of non-ideal situations (hunger, starvation) or life choices (gluttony).

As such, this evaluative stance taken towards consumption in *PP*, as evidenced in the collocates for *drink* and the poem's key lexis more generally, reflects a degree of negation difficult to ignore. The collective negativity conveyed in terms such as *hunger, starving,* and *beg(ging)* suggests a specific evaluative approach ascribed to the topic by

Langland in its framing, one which would presumably resonate with his late medieval communal readership at large. None of the consumption-based KWs in Figure 5.2 convey overt positivity with the possible exemption of *moderation* and *temperancie* but these alone do not change the cumulatively negative linguistic landscape evidenced by looking at these KWs collectively as a group.

The negative semantic prosody evidenced in the poem's consumption KWs can be further explored through evaluative language studies. Thompson and Alba-Juez (2014: 9-14) provide a summary of the terminology used in recent studies to address evaluation, but it is Hunston and Thompson's (2000: 5) definition which is used here based on its comprehensiveness in addressing the core tenets of evaluation:

> [Evaluation is] the broad cover term for the expression of the speaker's or writer's attitude or stance towards, view point on, or feelings about the entities or propositions that he or she is talking about. That attitude may relate to certainty or obligation or desirability or any of a number of other sets of values… [expressing] a user-orientation in terms of the two perspectives mentioned earlier (it is the user who evaluates), but it also allows us to talk about the values ascribed to the entities or propositions which are *evaluated*. (Hunston & Thompson 2000: 5, original emphasis)

This conceptualisation of evaluation takes the approach that evaluative language is comprised of two components, evaluation and stance. In making this distinction, Hunston (2011: 51) argues that evaluation ascribes value to an entity while stance, or 'interactivity' more generally, refers to textual indicators that 'a human being, the writer, is communicating with another human being, the reader'. This is an important differentiation, for *stance* in this sense would involve analysis focusing on the text as Langland's medium of communication to his readership, with study of *evaluation* oriented around addressing the assessment found within the text as a predominantly text-centred phenomenon. The non-mutual exclusivity of these aspects

is also addressed by Hunston (2011: 51) who acknowledges the overlap existing between the two in having 'a text that is highly interactive but does not contain much evaluation, or one where indications of interactivity are largely confined to evaluation'. From amongst these two options, it is not difficult to see *PP* – and, arguably, many classical literary text-types – as examples of the latter.

The degree of abstractness typical in discussing evaluative language is, perhaps more than anything else, premised on the inability to ascribe a set list of grammatical or lexical forms as the sum of ways in which evaluation can be linguistically expressed. While adjectives and adverbs are typically associated with evaluation through processes such as characterisation, 'this does not mean that every adjective and adverb marks evaluation or that all evaluation can be identified in this way' (Hunston 2011: 3).[59] In addition to (or perhaps as a result of) this absence of correlating form-to-meaning, evaluation is often not overt but rather implicit and expressed cumulatively as the meaning generated from several recurring textual features; this includes instances of phrasal evaluation such as Louw and Milojkovic's (2016: 345) example of *bent on* where the expression itself is evaluative but its individual components in isolation are not.

Corpus approaches have proven useful to date in accounting for the pragmatic and stylistic significance of evaluation (if not its initial identification) in historical Englishes, typically in a combined cognitive-corpus approach. Fabizak & Hebda (2010), for example, conduct a corpus-informed analysis of *pride* in OE as a metaphor with changing evaluative polarity in the conflict of pagan and Christian ideologies, while Diaz-Vera (2014) looks at the evaluation embedded in emotions of *guilt* and *shame* in

[59] See Hunston & Thompson (2000: 14) for additional examples of lexis that is chiefly evaluative in function and meaning, including nouns and verbs in addition to adjectival and adverbial forms.

OE culture. Claridge (2012) takes a more feature-oriented approach in using the Helsinki corpora to show how hyperbole operates as a stylistic magnifier of positive or negative emotional evaluation in OE writing, an approach correspondingly seen in Wallage's (2017) study of textual negation in ME, with evaluation peripherally recognised as a functional effect linked to negation in general.

In contrast to the cognitive grammars employed by most of the above corpus-assisted research on historical linguistic evaluation, the present study utilises the functional grammatical framework for classifying evaluation set out by Martin (2000) and later Martin and White (2005). One of the most comprehensive theoretical classifications created for analysis of evaluative language (Thompson & Alba-Juez 2014: 6), this functional approach operates within the tradition of sentiment analysis in mining the corpus to identify features of (and patterns contributing to) evaluative polarity. This choice of approach for the present study follows my already functional approach to style evidenced by my use of Mahlberg and McIntyre's (2001) classification framework for text function. It also follows modern corpus studies such as Whitelaw et al. (2005) and Bloom et al. (2007) in using their framework as the underlying basis for analysing evaluative language, albeit on a smaller evaluative scope as evidenced within a specialised literary context rather than of language generally.

Martin and White (2005) take an ideational approach to evaluation as a phenomenon where intersubjective stance can vary as social relations are negotiated through degrees of appraisal, involvement, and negotiation (see Figure 5.14). From amongst these three axes, Martin and White (2005: 34-8) focus on appraisal as the aspect most tangentially related to evaluation through its three sub-domains: attitude, engagement, and gradation:

> Attitude is concerned with our feelings, including emotional reactions, judgements of behaviour and evaluation of things. Engagement deals with sourcing attitudes and the play of voices around opinions in discourse. Graduation attends to grading phenomena whereby feelings are amplified and categories blurred (2005: 35).

From this definition we can start to see the aims of existing OE and ME scholarship on evaluation, with the attitudinal aspects of evaluation addressed in Wallage (2017) standing in contrast to the primarily graduation-based concern in the hyperbole featured in Claridge (2012). Interestingly, the metaphorical significance attached to concepts of pride, shame, and guilt explored in Fabizak & Hebda (2010) and Diaz-Vera (2014) address both attitude and engagement domains simultaneously, first in the identification of the emotional responses surrounding the use of these lexemes (attitude) followed by consideration of the socio-cultural reasons behind these responses (engagement).

Figure 5.14: Chart from Martin and White (2005: 38) outlining the sub-domains (and the respective sub-groupings) of appraisal used in categorising types of evaluative language.

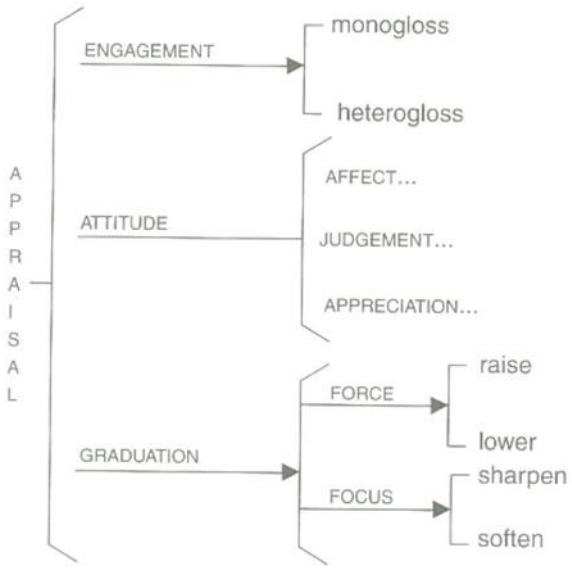


Figure 1.18 An overview of appraisal resources

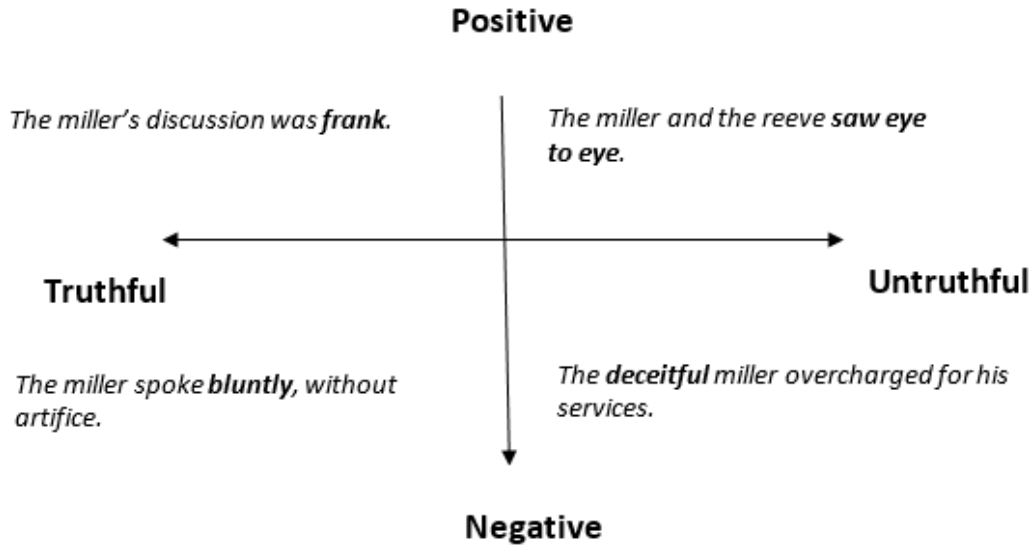
The present interest in using Martin and White's (2005) framework is to focus on the methods by which attitude is embedded in the poetic use of *drink* in *PPCorp*, with the KW acting as a representative example for the other negatively evaluative key lexis observed in the poem. This does not preclude observations about engagement or graduation but rather prioritises attitude-based appraisal in its focus on evaluation through JUDGEMENT, APPRECIATION, and AFFECT (see Figure 5.14). In brief, AFFECT relates to the presentation of feelings and JUDGEMENT reveals general attitudes towards behaviour, with APPRECIATION focusing on attributed valuation within a given field, area of interest, or larger social demographic (Martin & White 2005: 42-3). All three aspects of attitude can be conveyed positively or negatively, with the significance of negative AFFECT being completely different than negative APPRECIATION; it is also possible for an evaluation to constitute a combination of two or more of these attitude-based sub-categories. Each of these aspects of attitude can be further explored by looking at the values specific to each type of evaluation; AFFECT, for example, can be evaluatively measured along parameters of (*un*)*desirability*, (*in*)*security*, and (*dis*)*satisfaction*, emotional values which do not translate into the more behaviour-oriented trends characteristic of JUDGEMENT, as shown below in Table 5.15.

Table 5.15: The parameters or 'values' associated with evaluative judgement, affect, and appreciation, according to Martin and White (2005: 53; 56; 71).

TYPE OF ATTITUDINAL EVALUATION	VALUATION
AFFECT (EMOTION)	desire (un)happiness (in)security (dis)satisfaction
JUDGEMENT (BEHAVIOUR)	normality (how special) capacity (how capable) tenacity (how dependable) veracity (how honest) propriety (how beyond reproach)
APPRECIATION (VALUE)	reaction (nature of impact/quality) composition (how balanced/complex) valuation (how worthwhile)

These parameters present a method of both identifying and classifying instances of textual evaluation, with valuation as one axis and its positive or negative polarity as the second. As demonstrated below in Figure 5.16, these evaluations can manifest themselves in a range of forms, including adjectives, adverbs, and some non-evaluatively explicit phrasal language. When exploring representations of *veracity* terms such as *blunt* and *deceitful* are both fundamentally negative but *blunt* contains meaning linked to truthfulness not found in instances involving *deceit*. Similarly, being *frank* is generally viewed as a positive attribute of candid honesty, whereas the positive meaning in *see eye to eye* (or, being in agreement with someone) is offset by external speculation that agreement between a miller and reeve – particularly in a medieval literary context – could well imply some sort of dishonest fiscal arrangement enacted for mutual gain.

Figure 5.16: Examples of positive and negative representation of *veracity* (evaluation of truthfulness) as a manifestation of JUDGEMENT, using lexical and phraseological examples drawn from different grammatical categories



Integrating Martin and White's (2005) framework into the current investigation uses the negative polarity evidenced in *drink*'s collocates as a point of departure for establishing the underlying valuations by which this evaluation is achieved. In other words, the aim of using Martin and White (2005) is to determine the leading evaluative parameters of appraisal contributing to representations of drink(ing) in the text. This goal is achieved through concordance analysis of the 36 occurrences of *drink* in *PPCorp*; I use the numbering system of the concordance lines set out in Figure 5.13, but return to the original unannotated text to conduct this evaluative analysis. Each concordance line is first categorised by valuation (e.g. *satisfaction*, *veracity*, *capacity*), which can be used to not only pinpoint the most common parameter of evaluation framing drink(ing) in *PP* but also the most characteristic type of attudinal evaluation for that KW (AFFECT, JUDGEMENT, APPRECIATION). This approach, in addition to offering greater insight into the evaluative framining of drink(ing) in the poem, provides a

transferrable framework enabling further comparative analysis of the use of *drink* in *PP* to the other texts in the reference corpus, as done in 5.4.5.

5.4.3 Types of appraisal

Table 5.17 below contains an annotated chart listing the different types of evaluative expression taking place in each of the 36 instances of *drink* in *PPCorp*. Its columns (moving left to right) signify the keyword's use in the text as a noun or verb, the type of attitudinal evaluation taking place, and the thing being appraised in each textual instance. When classifying each use as a form of AFFECT, JUDGEMENT, or APPRECIATION, the evaluative polarity (positivity/negativity) conveyed in the text in each instance is signified by a (+) or (-) sign.

Table 5.17: Attitude-based appraisal observable in the 36 occurrences of *drink* in *PPCorp*; examples align with L1 ordering in the concordance lines present in Figure 5.12, with the layout modelled on the appraisal analysis conventions in Martin & White (2005: 73-4).

N.	KW function	Appraising items	Affect	Judgement	Appreciation	Appraised
1	Drinking	*feble*	(-) satisfaction			Diluted ale
4	Drinking	*bote* 'cure'		(+) propriety		Salvation from Christ (fig.)
15	Drinking	*riht rype* 'ripe'	(+) satisfaction			Good wine
19	Drink	*fynde* 'provide'		(+) normality		Sustenance provided through work
29	Drink	*þat doth good*		(+) propriety		Acceptable approach to drinking
31	Drinking	*nat out of tyme*		(-) propriety		Drinking before noon
32	Drinking	*frete* 'break'		(-) propriety		Break fastdays (drink) before noon
36	Drinking	*nat ouerdelycatly no to depe*		(-) propriety		Immoderate alcohol consumption
37	Drinking	*broke*		(-) normality/propriety		Drinking from a brook
38	Drinking	*þat thow madest*		(-) propriety		Self-made situations (fig.)

39	Drink	*drede delitable* 'dread delicious'	(-) inclination			Wariness of alcohol
40	Drink	*dryue*		(-) capacity		Drinking to drive away emotions
41	Drink	*drosenes and dregges* 'dregs'	(-) satisfaction			Drink for beggars
42	Drink	*nat care for*	(+) security			Worry about basic sustenance
43	Drink	*likerous* 'liking of'	(+) inclination		(+) appreciation	Excessive fondness for drink
44	Drinking	*deth to lette* 'delay'		(+) propriety		Drink to delay death (fig.)
45	Drink	*prest* 'pressed' *to plese*	(-) security			Poor pleasing the rich for food/drink
46	Drink	*out of mesure*		(-) propriety		Overconsumption
47	Drinking	*but with þe doke*		(+) propriety/(-) normality		Drinking water (fig.)
48	Drink	*leseth*		(-) propriety		Losing grace through drinking
49	Drinking	*drat* 'fear' *no synne*		(+) capacity		Drinking elim. fear of sinning (fig.)
50	Drink	*loue*		(+) propriety		Love as type of drink (fig.)
51	Drink	*helpe*		(+) propriety		Church helping poor with needs
52	Drinking	*wyne ofte*		(+) propriety		Poor not drinking wine often
53	Drink	*swete*	(+) satisfaction			Sweet drink

54	Drink	*deyede* 'died'		(+) propriety		Death for love (drink type) (fig.)
55	Drink	*deth and depe helle*		(-) propriety		Drink as death (fig.)
56	Drinking	*goed ale*	(+) satisfaction			Drinking quality products
57	Drinking	*alday*		(-) propriety		Drinking all day
58	Drinking	*no dede to do but*		(-) propriety		Drinking as form of laziness
59	Drinking	*druie* 'drain the pot' *and depe*		(-) propriety		Excessive consumption
60	Drink	*drow vs* 'drew us'		(+) normality		Drink as part of toasting process
61	Drinking	*noon halpeny ale*	(-) satisfaction			Wasters
62	Drinking	*deep clergyse* 'learning'		(-) propriety		Christ. devotion over learning (fig.)
63	Drinking	*late at euen* 'evening'		(-) normality/propriety		Drinking late at night
64	Drinking	*ar thow dyne*		(-) propriety		Drink before eating

In looking at Table 5.17 a distinctive evaluative pattern emerges surrounding drink(ing). The vast majority of instances of *drink* (or, in the original text, *drynke*) are shown to involve at least one type of valuation associated with JUDGEMENT, with several addressing aspects of AFFECT. The relative dominance of JUDGEMENT is particularly visible in Table 5.17 and alone accounts for 26 out of the total 36 uses of the KW in *PP*. The comparative absence of APPRECIATION relative to exemplified AFFECT and JUDGEMENT characterises the textual approach taken to drinking in the poem as one that largely avoids articulations of value (e.g., *how worthwhile*) in preference for emotion- and behaviour-oriented representations. The current discussion focuses on the examples of AFFECT and JUDGEMENT in *PP*, with particular emphasis on a) the most frequent parameters within these types of appraisal and b) examples including multiple valuations simultaneously.

5.4.3.1 Affect

Instances featuring AFFECT predominantly occur along evaluative parameters of personal *satisfaction*, typically involving specific types of beverages. Positive and negative evaluation along these lines manifests itself adjectivally based on the quality of the drink, as seen with references to 'ripe wine' (15), 'good ale' (56) and more generally with 'sweet drink' (53). Negative representations of *satisfaction* similarly revolve around certain beverages like 'feeble ale' (1), but also link the inferior quality of these drinks to the individuals conventionally linked to their consumption, with dregs associated with beggars (41) and half-penny ale to wasters, or those not gainfully employed in honest work (61).

Evaluative AFFECT is also manifested in *security* as a disposition of emotional disquietude (Martin & White 2005: 50) or concern surrounding drinking. This is shown to relate to worries regarding the individual safety and overall well-being experienced by persons as either a natural need to eat and drink to live in general (42), or an obedience to the wishes of the rich in order to acquire said basic sustenance (45). These examples represent the poor as a targeted demographic, with (42) favourably observing how the church sees to basic life necessities (food, drink, clothes) so that the poor do not have to be concerned about the means of maintaining their existence, as is required in (45). AFFECT can furthermore be seen in the poem through *inclination* as both disposition and behaviour (Martin & White 2005: 48-51), with examples of these universally negated in the poem (39, 43). The imperative command issued in (39) warning to 'dread delicious drink' conveys AFFECT as a mental process highlighting the underlying wariness that should be maintained towards drinking as a practice, a type of concerned awareness manifesting as a gradated version of fear (Martin & White 2005: 48) in approaching drink; this wariness is in this instance lexicalised in the term *drede.* By contrast, (43) offers an example of an affective behaviour in the *inclination* of the Biblical figure Lot to drink, with some scripturally problematic results.

5.4.3.2 Appreciation

Some uses of *drink* involve more than one aspect of attitude-based appraisal, as exemplified in the text's sole instance of APPRECIATION in Table 5.17. While APPRECIATION is not dissimilar to the *satisfaction* discussed above with AFFECT, APPRECIATION of drink would require a personal reaction to its quality (e.g., 'this ale is

_____'), rather than observance of the quality itself. In the case of (43), Lot's predilection for drink can be interpreted not only as an *inclination* of behaviour but also as a *reaction* to the quality of the drink being consumed (Martin & White 2005: 56). His approval of drink is not explicit but rather inferred from the already positive affective evaluation revealed in the text by the use of the phrase *likerous drynke* 'excessive fondness of/for drink' as a type of *inclination*:

(iii) Loot in his lyue thorw **likerous drynke**
Wykkede wroghte and wrathed god almyhty
In his droknkenesse aday his doughteres he dighte
And lay by hem bothe as þe boke telleth (43)

*Lot in his life, through his **excessive fondness for drink***
Worked wickedly and angered the almighty God
In his drunkenness one day he topped his daughters
And lay with them both, as the book tells us. (1:25-8)

In this sense, Lot appears to like drink (i.e., appreciate it) to the point where it becomes an emotionally-driven gluttonous inclination that characterises his subsequent decisions and actions.

5.4.3.3 Judgement

As the most frequently recurring type of evaluation associated with drink(ing) in the poem, JUDGEMENT (like AFFECT) is expressed across several evaluative parameters including *normality*, *capacity*, and *propriety* (Martin & White 2005: 53). Instances of *normality*-based judgement, as it sounds, conducts evaluation along the boundaries of conventional expectation. These normative assessments can be seen in instances appearing alongside another parameter of JUDGEMENT, like the *propriety* visible in (37) and (47). In the case of (37), the action of drinking from a brook is negatively

evaluative from the perspective of both *normality* (how often does someone drink from a brook?) as well as *propriety* (not typically an acceptable action, especially for adults). Example (47) also features this dual appraisal through *normality* and *propriety* in the figurative expression *drynke but with þe doke* 'drink but with the duck', but with different evaluative polarities; drinking water is good from a *propriety* perspective in that it prevents many of the problematic behaviours associated with consuming alcohol, but the manner of drinking this water is unusual, as water was rarely consumed in large quantities in late medieval English culture (Bennett 1996). Its status as a non-*normative* action becomes even more negatively characterised if interpreted literally as drinking directly from a pond.

Within JUDGEMENT, these instances of *normality* represent textual examples of expressed social esteem. Despite the positive connotations associated with 'esteem' the term in an evaluative context refers to social opinion divisible along set polarities. Positive expressions of *normality* are indicative of actions, decisions, behaviours, or opinions garnering social admiration, while those not aligning with conventional cultural values tend to generate social criticism (Martin & White 2005: 52; reproduced in Figure 5.18 below). The positive *propriety* in (47) thus reflects a general socially-maintained approval of the steps being taken by the character Lechery to stay out of mischief, while the negative *normality* reveals the underlying social criticism involving the manner in which this consumption of water takes place.

Table 5.18: Social esteem and sanction and the effects produced by each in judgement-based evaluation (based on Martin & White 2005: 52-3)

	Positive (+)	Negative (-)
Social esteem (*normality, capacity, tenacity*)	Admiration	Criticism
Social sanction (*veracity, propriety*)	Praise	Condemnation

Social esteem can also be conveyed through evaluation of *capacity*, or the ability to do something (Martin & White 2005: 55). This *capacity* is conveyed as criticism in (40) through the professed inability of alcohol to drive away or otherwise eradicate feelings of envy or general ill-will, and also features in discussion of the ability of drink to drive away fear of sin (49). This ability, while positive from a strictly *capacity*-based standpoint, functions as a criticism from a social viewpoint; the effect itself is not viewed as largely positive (i.e., the ability of drink to remove fear of committing wrongdoing), but its textual use is evaluatively positive in that (49) references drink's abilities to produce or otherwise achieve something, rather than its inability to do so.

One thing worth briefly noting here is the comparative absence of valuations of *tenacity* and *veracity* in *PP*. This could be based on the choice of topic in looking at drink(ing), as well as the poetic format of the text itself; discussions regarding the drive or resolution to drink (*tenacity*) and the truthfulness or authenticity of accounts of drinking (*veracity*) would be more typical in text-types associated with self-help (i.e., pamphlets), rather than in narrative poetry. Instead, the valuations underlying representation of *drink* in *PP* as a poetic text appear to be based to some extent on socially-driven valuations of *normality* and *capacity*, but primarily driven by evaluative JUDGEMENT based on ethical *propriety*.

5.4.4 Judging drink(ing) through propriety

Within this framework of negative JUDGEMENT, Table 5.17 reveals *propriety* to be the most common method through which drink(ing) is evaluated in *PP*. Unlike instances of social esteem caused by valuations like *normality* and *capacity*, *propriety* represents a form of social sanction, with the difference between esteem and sanction set out in more detail by Martin and White (2005: 52):

> Social esteem tends to be policed in the oral culture, through chat, gossip, jokes and stories of various kinds – with humour often having a critical role to play (Eggins & Slade 1997). Sharing values in this area is critical to the formation of social networks (family, friends, colleagues, etc.). Social sanction on the other hand is more often codified in writing, as edicts, decrees, rules, regulations and laws about how to behave as surveilled by church and state – with penalties and punishments as levers against those not complying with the code. Sharing values in this area underpins civic duty and religious observances.

The overwhelming textual emphasis on examples of social sanction through *propriety* visible in Table 5.17 orients poetic concentration on *drink* (and to a larger extent, consumption) as a fundamentally moral subject. Sources of moral evaluation are typically linked to specific authoritative institutions that disseminate, maintain, and monitor their social use; within a medieval context, such institutions were most frequently associated with the church and its teachings (Gruenler 2017: 221). While not acting as a codification of rules regulating practices of drinking, the poem does make recurring use of (and reference to) biblical material and doctrine produced by the church in its intertextual allusions and direct biblical quotations. This adds an additional religious aspect to the topic beyond the everyday practicalities of eating and drinking to stay alive by establishing ways in which consumption is *meant* to occur. This connection between *propriety* and morality thus situates morality as an evaluative

attitude on the level of text, with the majority of examples of *drink* exhibiting aspects of this distinctive moral attitude. In other words, the majority of the representations of drink(ing) in *PP* address 'how far beyond [moral] reproach' (Martin & White 2005: 53; clarification added) individual or communial behaviours are along a cline of ethical assessment.

Expressions of social sanction, as linked to eating and drinking, is a theme recurringly identified in late medieval literary criticism. Studies of the literary representations of women brewers (Bennett 1996) alongside depictions of retailers in contemporary legal and literary texts (Carrell 2006 and Jones 2011, respectively) reveal common dimensions through which persons were evaluated in a 14th and 15th century context. These dimensions include, but are not limited to:

(a) Physical attributes
(b) Actions or habitual behaviours
(c) Contextual surroundings (setting)
(d) Close associates
(e) Sexuality (or lack thereof)

These evaulative dimensions (a-e) can all be shown to be linked to *propriety*-based signficance in a late medieval context. For example, the description of the cook in *CT*'s *General Prologue* as a slovenly, dirty, and diseased individual (a) contains evaluation along parameters of *propriety* (*how far beyond moral reproach?*) for, as Carrell (2006: 187-8) observes, physical dirt and illness in particular were often culturally equated with moral uncleanliness. Thus, this representation of the cook not only generates social criticism based on his unusually ugly, dirty, or otherwise disease-ridden appearance, but also creates additional moral sanction condemning him for preparing and distributing food despite being potentially unfit to do so based on bodily visible indicators of poor health and/or hygiene. This preoccupation with morality and

social monitoring of it, Jones (2011: 17) notes, was part of 'an imaginative preoccupation with London' and 'the literary formation of the commercial marketplace as the moral challenge of Langland's age'.

The morality-based sanction of drink(ing) in *PP* particularly focuses on consumptive moderation. Moderation has already been identified as a topic of interest in the poem through the other consumption-based keywords generated for *PPCorp* by states such as *moderation* and *temperancie*, and more implicitly in morally-charged terms denoting temporal events linked to occasions of voluntary consumptive moderation (i.e., *lent*). The evaluative 'appropriateness', or civic *propriety*, associated with moderation features in the use of *drink* as both a concept and a practice; this can be seen temporally in statements evaluating drinking before noon (32, 31), late in the evening (63), or even *alday* (57), as well as through manner in discussions of drinking before eating (64). There is also evidence of evaluated moderation via amount in drinking too much (36, 46), and even through predilection, or excessive gluttonous fondness, for drink (43). This characterisation of moderate drink(ing) is predominantly negative in its grammatical framing, particularly within exhortations to *drynke nat* (36, 31), or to *ne [drynke] ofte* (52) as actions to avoid, rather than behaviours in which to aspire. This approach to alcohol consumption mirrors – and, to a degree, is shown to primarily fuel – the broadly negative approach to *drink* initially observed in 5.4.1.

In looking at Table 5.17 it is clear that not all *propriety*-derived evaluations of moderation in the poem are negative; however, negativity appears to be the dominant approach. As with social esteem, the polarity characterising moderation determines its pragmatic meaning, with positive evaluations of *propriety* leading to social praise while negative appraisal produces outright social condemnation of the person, action, or situation being evaluated. The differences between these approaches to moderation,

as illustrated in the examples found within *PP*, are further parsed below in 5.4.4.1 and 5.4.4.2.

5.4.4.1 Social praise

The praise, or positive social sanction, attributed to moderation in *PP* is generally oriented around descriptions of the poor. This type of praise is based on *propriety*-derived representations of the poor living a good, moral life (51, 52). From a modern perspective it seems unusual to attribute praise to the condition of being poor, a socio-fiscal situation which is prototypically viewed with criticism today based on the negative perception of poverty with a standard of living deemed problematic along a valuation of *security*. By contrast, the poem (and to an extent, a late 14th century worldview more generally) takes a different approach to poverty in emphasising the *propriety*-based effects of this condition, wherein having little inadvertently enables a praiseworthy lifestyle through not having the money to be gluttonous or drink excessively. Sourced from ecclesiastical doctrine, this approach draws a distinctive line between those who work hard and are still poor, and individuals experiencing poverty through a decision not to work, or the 'wasters' alluded to in (61). The praise conveying moral *propriety* to the former demographic is particularly visible in observations made on the drinking habits of the poor (52). The fact that late medieval poverty in this context was not typically voluntary but a decision based on situational conditions is largely overlooked in broadly ascribing positivity to both those who have control over their decisions and choose to live moderately (i.e., morally), and those who do not and live a moderate lifestyle anyway out of basic necessity. To this end, it

is worth observing that the textual emphasis on moderation is not often concerned with repercussions to an individual's health, but rather the moral significance attached to adherence to practices of moderation (see 29).

Additional examples of social praise occur through figurative allusions to religious doctrine. The drinking from the breast of Christ for redemptive salvation described in (4), for example, presents an abnormal image realistically but portrays a process well-represented in biblical doctrine wherein drinking from an evaluatively positive source transcends previous moral wrongdoing and operates as a praiseworthy action to take (see also 44). The recurring portrayal of love as a drink (50, 54) is particularly interesting to find in the context of evaluative *propriety*; love typically operates as an emotion experienced by someone (AFFECT) or a reaction to something (APPRECIATION) but the concept is used here to depict good moral conduct, where someone subsisting on love stands above any type of social condemnation (JUDGEMENT).

5.4.4.2 Social condemnation

By contrast, the poem's *propriety*-based condemnation focuses primarily on the consumption of alcohol; this holistically represents the most frequent type of evaluation that occurs with *drink* in *PPCorp*. As mentioned previously, condemnation adopts a distinctively negative tone in its recurring message that all efforts should be made to avoid inebriation (31-32, 36, 46, 48, 57- 59, 63-64). These evaluations are conveyed in the poem through imperative statements either exhorting the avoidance of alcoholic beverages through effects including the loss of one's grace (48) and signs

of laziness (58), or through advised moderation when consuming said drinks (46, 59). The foundation for these imperatives can be found based on behavioural observation such as people drinking 'out of measure' (46) and 'breaking fastdays before noon' (32), as well as through learned assertions exemplified in the statement *and drynke nat ouerdelycatly no to depe neyther* (36). Some of these discussions of avoiding inebriation are positively evaluative as already observed in (47) and (52) but the preponderance of textual examples feature negative proprietary evaluation. Textual instances containing both positive and negative evaluation are particularly powerful, as seen in Holy Church's discussion with the narrator regarding the three essentials required by people in passus 1:

(iv) The firste is fode and vesture þe seconde
And **drynke þat doth the good** and **drynke nat out of tyme** (29, 31)

The first is food and clothing the second
*And **drink that does you good**, but **don't drink out of time*** (1:23-4)

In this particular instance, (29) features a positive statement referring to drink as not only an acceptable but vital part of life's sustenance, while (31) contains an explicitly condemning statement cautioning moderation in alcohol-based drinking practices.

Much of this social condemnation peripherally brushes other valuations of JUDGEMENT, with some more prototypically *propriety*-based than others. Taking (32) as an example, it would be *normality* to state that drinking typically doesn't occur before noon, but this instance is a case of evaluation through *propriety* in that it acknowledges this norm as a given state of existence, with evaluative judgement based on the explicit non-adherence to this standard of expected behaviour taking place in the events within the text (in this case, during the festivities surrounding Meed's unsanctioned marriage). This sense of moderation conveyed through parameters of

propriety is more clearly evidenced in textual examples such as (46) in the phrase *out of mesure*, an evaluation applying to consumption of both food (*mete*) and drink when the narrator questions Reason's choice to rule over men than other types of beasts in passus 13:

(v) y se non so ofte forfeten sothly so mankynde
In mete **out of mesure and mony tymes in drynke**
In wommen in wedes and in wordes bothe
They ouerdoen hit day and nyhte and so doth nat oþer bestes
They reule hem al by resoun ac renkes ful fewe (46)

I see none as often forfeit, for sure, as mankind
In meat ***out of measure and many times in drink***
In women in both clothes and in words
They overdo it day and night as other beasts do not
They [the beasts] rule themselves by reason but few men do (13:193-7; clarification added)

This representation of the moderation here as practiced by people in general is particularly critical in its appraisal, suggesting that the moderation of consumption in general (not just dietary but material in the reference to clothing) is a fundamental problem. This concern also manifests itself beyond literal moral monitoring of drink(ing) into figurative depictions. One of the most striking examples of this appears in (55), which metaphorically associates drinking with both death and hell simultaneously (see ii).

5.4.5 PPCorp vs. MEPoetLitCorp

The detailed concordance analysis conducted in 5.4.3 and 5.4.4 is useful in that it provides an evaluative profile of *drink* within *PP*. This profile reveals that *drink* is a) prototypically represented via parameters of JUDGEMENT, and b) within JUDGEMENT evaluation is textually realised through valuations of *capacity* (how capable), *normality*

(how typical), and most notably *propriety* (how beyond reproach). The linguistic negation used within these valuations has contributed to the production of effects such as social praise and condemnation, with the praise addressing aspects of consumption viewed negatively in a modern context if not to its medieval readership (e.g., the praise of poverty in the working poor).

This profile of *drink* in *PP* can thus be compared to the profile of the term's use in the poetic reference corpus. Such a comparison would identify whether the approach in the poem is reflective of the general approach evidenced in late medieval poetry in general or one unique to *PP*, thus providing the grounds to identify these processes as being stylistic or register-based. To effectuate this comparison, concordance lists were generated for *drink* from *MEPoetLitCorp* and its results similarly examined according to Martin and White's (2005) evaluative framework. While there are too many occurrences of the term in the reference corpus to illustrate the level of detailed analysis demonstrated for *PPCorp* in Table 5.17, I use illustrative examples selected from *ARCorp* and *MVCorp* to separately highlight the comparative similarities and differences evidenced between each of the (sub)corpora.

5.4.5.1 MVCorp

The 119 occurrences of *drink* in *MVCorp* (see Figure 5.19 on CD) reveal some broad similarities in how drink(ing) is approached in the reference corpus' metrical texts. As seen in *PPCorp*, *MVCorp* contains instances of emotionally-driven AFFECT, exemplified in the *inclination* to drink wine but choosing to drink water discussed in *CA*:

(vi) Thus was he turned fro delights:
The wine which he was wont to **drink**
He took instead of the wells brink
Or of the pet or of the swamp
It thought him then good enough (100, *CA*)

AFFECT can also be seen in the negated valuation of *happiness*, which is used in the corpus to measure an emotional response to drink through love as a type of woe to be experienced (105, *CA*), with woe existing as a gradated version of sadness (Martin & White 2005: 51). This thematic connection between love and drink(ing) appears only peripherally in *PP,* a topical association also featured in *CA* through APPRECIATION in evaluative comparisons of love as being a 'bittersweet thirst' (59, *CA*) as a reaction to perceived quality of experience (Martin & White 2005: 56). Instances of AFFECT are also shown to operate similarly to the *(dis)satisfaction* evidenced in *PPCorp* through characterising descriptions of drink as a beverage being 'sweet' (60 and 107, *CA*), 'strong' (103, *GP*), 'rich' (88, *PT*), 'good' (46, *ManT, P*), or 'bitter' (40, *TC*).

These instances of (or opportunities for further reflection on) social esteem in *MVCorp* also include some behaviour-based evaluations of JUDGEMENT through *tenacity* and *normality*. The presence of *tenacity* in *PardT, P* lexicalised in the use of the modal *will* as a modulation of inclination (Martin & White 2005: 55) is striking, given the absence of dependability linked to *drynke* in *PP*:

(vii) I will have money, will, choose, and wheat,
Al were it given of the povrest page,
Or of the povrest widwe in a village,
Al should her children starve for famine.
Nay! I will **drink** liquor of the vine,
And have a Holy wench in every town. (117, *PardT, P*)

MVCorp, as with *PPCorp*, also contains examples where *propriety* overlaps with other valuations in individual uses of *drink.* While usually occurring in the dual presence of

propriety and *normality* in *PPCorp*, this duality typically occurs in *MVCorp* via the merging of *tenacity* and *propriety*, as illustrated above in (vii). This *propriety* is observable in the underlying motivations for producing such an utterance; such a wilfully stated assertion would only need voicing aloud as a defence from the unspoken social condemnation surrounding habitual practices of heavy drinking and keeping company with women of questionable profession and perceived moral character.

What separates *MVCorp* from *PPCorp* in the evaluative presentation of drink(ing) is the proportional difference in the amount of *propriety* evidenced in each corpus. *PPCorp* is fundamentally characterised in its *propriety*-based presentation of *drink* for the purposes of morally-governed moderation, a stance drawn from its closeness in tone, scope, and topical interest in concerns raised (and beliefs maintained) by contemporary religious establishments. However, much of the evaluation in *MVCorp* avoids this key moral component in its discussion of drink(ing) in preference for the more factually-oriented approaches to the topic as being *normative* behaviour. This is observable in the markers of usuality (Martin & White 2005: 54-5) featuring in discussions observing the departure of reason and wit associated with drinking alcohol referenced in *ParsT* (viii) and *CA* (ix) (see also 79, *CA*).

(viii) For dronkenesse is very sepulture
Of mans wit and his discrecioun.
In whom that **drink** hath dominacioun,
He can no counsel keep, it is no dread. (92, *ParsT*)

(ix) That **drink** which makth the heart brenne,
And doth the wit about run (91, *CA*)

Both (viii) and (ix) contain descriptively-based factual observations about the repercussions of immoderate alcohol consumption. What is notable about these uses

of *drink* compared to examples from *PPCorp* like (iv) and (v) is the absence of lexicalised *propriety* evidenced grammatically through negative articles and modals stating that an action 'should not' occur under a given set of conditions, if at all. This absence of morality-based JUDGEMENT in approaching drink(ing) in this section of the reference corpus stands in sharp contrast to what is characterisitically expected in *PPCorp*.

This is not to say that propriety does not feature in *MVCorp*, but rather that a more balanced landscape of evaluative types characterise this reference sub-corpus relative to *PP*. Clear combinations of *normality* and *propriety* are visible in voiced opinions such as the link between ale brewers and drinking problems raised in *CA* (115), while examples of overt *propriety* can be seen in imperative exhortations for temperate drinking practices (65, *SumT*; see also 77, 106, and 119 in *RR*) and arguments made on moral grounds for the public to forsake and otherwise ostracise those characterised by their drinking habits (111, *CA*). The corpus also briefly addresses topics of more focused interest in *PP*, including the ethical relationship between hard work and sustenance (113, *PT*) and the moral stance that the church should supply food, drink, and clothing to the working poor (35, *CA*).

However, the largely negative *propriety*-based JUDGEMENT characterising *PP* stands in stark contrast to *MVCorp*'s focus on positive appraisal through a combination of AFFECT and JUDGEMENT. This evaluation predominantly addresses the topic of love as both an emotion (to love) as well as an experience (being in love), and how these are both described in terms of drink(ing). Instances of love as an emotion are manifested in feelings equivocated to being drunk, with AFFECT-based evaluation operating along parameters of (*un*)*happiness* through the characterisation of love-inspired *woe* as a type of drink (x). These evaluations are also presented as the

(*dis*)*satisfaction* manifested in the dispositional mix of interest and pleasure (Martin & White 2005: 51) present in the framing of love being like 'being drunk on someone' (xi). Contrastively, experiences of love can additionally be conveyed in the negated JUDGEMENT-based *normality* of being unable to perform normal life functions as a result of said emotions (xii).

(x) For love is yet the most stormy life,
Right of him-self, that ever was bigonne;
For ever some mistrust, or foolish strife,
There is in love, some cloud is over the sun:
Ther-to we wrecched women no-thing ability,
When us is wo, but weep and sit and think;
Our wretch is this, our own wo to **drink**. (105, *TC*)

(xi) How Tristram was of love drunk
With Inflamed Ysolde, when they drunk
The **drink** which Brangwein hem betok,
Er that king Marc his Eem hire took
To wyve, as it was after know. (95, *TC*)

(xii) When that Arcite to Thebes come was,
Full oft a day he swelte and said alas,'
For seen his lady shall he never-mo.
And shortly to concluden al his wo,
So much sorrow had never creature
That is, or shall, while that the world may continue.
His sleep, his mete, his **drink** is him biraft,
That lene he wax, and dry as is a shaft. (54, *KT*)

All three examples constitute occasions or situations involving involuntary non-consumption, albeit of a slightly different nature than the ascetic fasting encouraged in *PP*. They all similarly represent instances facilitating social admiration for being so in love that typical emotional patterns and behaviours vanish entirely as a type of socially-condoned abnormal state of existence. The unusualness of the circumstances evidenced in these examples is consistently lexicalised in negated forms from an evaluative context (*woeful*, *drunk*, *bereft*) but the end result is viewed positively from a

social perspective. This positive response, as a type of social prase, recognises that only a person truly in love could exhibit such aggravated symptoms interfering with basic life processes of consumption, or be so governed by the experience that emotion dictates dietary practice. This approach also extends to metaphoric representations of the relationship between *drink, love*, and *lust* (63, *TC*), as well as the seemingly compounded inebriation caused by being drunk on love (47, *CA*) in the manner illustrated in (ix). The metaphoric blending is also present in discussions of love potions (66, *CA*) in addition to more unusual images in the descriptions of physiological lust being projected onto actual meat and drink (67, *CA*).

5.4.5.2 ARCorp

Unlike the more AFFECT-oriented appraisal characterising much of *MVCorp*, *ARCorp* contains a stronger focus on evaluative JUDGEMENT in discussion (and use) of *drink*. As illustrated below in Figure 5.20, individual occurrences feature positive examples of *reaction*-based APPRECIATION in *deliteable* 'delightful' *drink* (10, *WA*) and *god drink haden* 'good drink had' (11, *Cl*); however, these represent isolated instances in the corpus. Examples of negatively evaluative AFFECT appear with greater prominence through the correlation of processes of drinking being linked to *dole*, a term linked to emotional responses of misery or distress in *PTA* (1) and *WA* (23) as manifestations of (*un*)*happiness*. This negative portrayal can also be seen in references to the practice of 'delighting in drink' as a form of emotional solace or consolation (15, *DT*), an affective mental process wherein delight operates positively as a gradated version of pleasure-

Figure 5.20: The 28 total instances of *drink* in the annotated *ARCorp,* with entries ordered according to L1 position.

N	Concordance	File
1	e was puysonede to deed, Their he was deed of a **drink**, as dole es to here, That the curssede Cassa	PTA.txt
2	r viij monethis; Then shall he duale the with a **drink** at you full dere traistis." Then makis he mo	WA.txt
3	sen you me now prays, You shall be drechid of a **drink** a dra?te of vnsele, And all your ?eris ere ?	WA.txt
4	sen you me prays, Thu shall be drenchett with a **drink** a draught of unclean, And all the yherez þar	WA.txt
5	m rwed that he hem vp-rerde & ra?t hem food and **drink**, & oft that he hem vndyd, hard hit him the?t	Cl.txt
6	biliue, That the lude my?t have leave food and **drink** to cach; For vneþe was the noyce not a while	SG.txt
7	overed it fair; Sho bade me derelich draw and **drink** to herselven; And thus sho led me about	aMA.txt
8	and schone scharply they eat; That food and **drink** for ladies was luther to chew. Fellen dow	SJ.txt
9	he byholdes. He bad him dread never a dole but **drink** it Violently, And then the pistyll of the pr	WA.txt
10	that beris vs our fodis. Have we na deliteable **drink** of diverse wynes, But watir of a wale well o	WA.txt
11	land schewe. & they bigonne to be glad that god **drink** haden, & each mon with his match made him at	Cl.txt
12	nted, and Gawain is lenged, Bliþe bro?t was him **drink**, and they to bed ?eden with li?t. Sir Gawain	SG.txt
13	him to the tavern till he tayte worth; Doo him **drink** al night that he dry be at morrow, Sythen ke	WW.txt
14	orth, That bring ?e to ser bacus to bear up his **drink**. And Iupitir that Ioglour sum Iape bos hav	WA.txt
15	mekenes, that him most louis, Delight him in **Drink**, and other dere meytes, Set him to solace, a	DT.txt
16	e behaldis. He bad him dred never a dole and it **drink** violently, And then the pistill of the pri	WA.txt
17	the? men ben merry in mind when they have main **drink**, A ?ere ?ernes full ?erne, and ?eldez never	SG.txt
18	deed?, For mon-sworne, & men-scla?t, & to much **drink**, For þefte, & for þrepyng, vn-þonk may mon h	Cl.txt
19	That we spedly have spoken, there spared was no **drink**.' Then acoles he the kny?t and kisses him þr	SG.txt
20	st, That, be it one, be it other that they of **drink**, The gestis shall have the goblettis and the	WA.txt
21	the song of the bestis, They have no dainty of **drink**, nor of dere meites, But derkon euon down on	DT.txt
22	r end, And glotony is her god, with gloppyng of **drink**, And gladness in glees, and great joye ymake	PPC.txt
23	ouir the montayns merrily it schynes Or he mi?t **drink** any drop and then his dole nor?es. Then come	WA.txt
24	, To purchasen him his pursfull to pay for the **drink**. And, brother, when bernes ben full and holl	PPC.txt
25	. His folk fell to it fast and freschely they **drink**, But was na renke at to it ran at ever rase	WA.txt
26	broth, nor best upon lyve, Wine nor water to **drink** but wope of himself. Old scheldes and sch	SJ.txt
27	intez Then any burn upon bench had bro?t him to **drink** of wine. Gawan, that sate by the queen, To t	SG.txt
28	e blown the powdere: There is doust for your **drink**!' down to hem cries, And bid hem bible of	SJ.txt

based *satisfaction* (Martin & White 2005: 46, 51) but is interpreted negatively from a social perspective.

As seen with *PPCorp*, most of the JUDGEMENT appearing in *AVCorp*'s use of the keyword involves negative appraisal of *propriety*. Some instances are similar in approach to *PPCorp* in correlating the behaviours of individuals in a specific profession (in this case, friars) with practices of drinking as a form of verbal social condemnation evidenced in *PPC* (24). This similarity is also observable in the moderation-based themes emerging from references to *drink all night* (13, *WW*), *gloppyng* 'guzzling [greedily]' *of drink* (22, *PPC*), and *too much drink* (xiii):

(xiii) For courtesy, & colwarde & crooked deed?,
For mon-sworne, & men-scla?t, & **to much drink**,
For þefte, & for þrepyng, vn-þonk may mon have;
For roborrye, & riboudrye & resoune? vntrwe (18, *Cl*)[60]

In this particular instance, the immoderate consumption of drink is included amongst a list of socially unacceptable behaviours including ribaldry and unsound reasoning, as well as socially-sanctioned crimes such as theft and robbery.

A particularly distinctive method of *propriety*-based JUDGEMENT in *ARCorp* not seen in *PPCorp* is the correlation between drink(ing) and health-related individual demise. This correspondence of action to a resulting diminished existential state primarily centres on the consumption of poison, or 'unclean drink' (4, *WA*). As an ethical matter rather than one based on truthfulness or dependability – and operating far beyond expectations of typicality – the reproachfulness factor of using poison operates solidly within *propriety*, wherein the negative condemnation associated with its use (i.e., bad, immoral, unjust, even illegal) is lexicalised in references to the

[60] The ? present in this quotation (and visible in several other passages taken from *PPCorp* and *MEPoetLitCorp*) are meant to be yoghs (ȝ); this is representative of a larger problem *Wordsmith* encountered in processing this archaic letter, even with the corpora's text files saved in Unicode.

resulting death and *dole* (misfortune) caused by its administration (1, *PTA*). Example (xiv) contains another instance of this correlation in the metaphorical representation of death occurring as a result of drinking 'a draught of pain and misfortune'. This metaphoric substitution of a physical beverage for an experiential concept is also visible in the *doust* 'dust' *for your drink* (i.e., ashes) declaration in *SJ* (28).

(xiv) But neuer-þe-les I shall the neuen sen you me now prays,
You shall be drechid of a drink a dra?te of vnsele,
And all your ?eris ere ?eat ?are and your ?outhe fenyst (3, *WA*)

It is worth observing here that individual demise does not always refer to death in *ARCorp*, but can also refer to the demise of one's honour and integrity. This type of evaluative moral downfall is visible in *SG*, with the explanation of how an illicit kiss was received through the linked statement *there spared was no drink* (19).

To this end, the uses of *drink* in *PPCorp* appear to more closely align with those contained in *ARCorp* than *MVCorp*. This suggests fundamental differences in how each poetic text-type approaches drink and drinking, rather than a uniform method of representation characterising medieval poetry in general. The closer affinity of *PPCorp* to *ARCorp* can be attributed to the higher number of *Piers Plowman* tradition texts in the alliterative revival poetry (for example, *RR* and *PPC*) whose textual examples have prominently featured in discussion of negative JUDGEMENTS of propriety-based morality, and the themes of moderation arising from said social condemnation. It can also stem from the heavy moralistic overtones present in the *Pearl* poems (*SG*), with the negative approach to morality through religious exempla highlighting wrongdoing particularly characteristic of *Cl*.

This affinity of *ARCorp* to *PPCorp* in their evaluative approaches to drink(ing) can also be viewed as a characteristic feature of alliterative revival poetry in general as

a form that lends itself to engagement with more politicised topics like moderating alcohol intake. The frequent use of allegory within the *AR* texts assists in with this in providing a stylistic method of more figuratively representing drink(ing) in a way that enables free exercise of opinion without directly impugning identified individuals; in other words, allegory renders the produced social condemnation abstract, which produces a level of universality in its interpretation. The most direct implications emerging from these works avoid recognisable contemporary persons in sanctioning either historical figures (the biblical Lot in *Cl* and *PP*, or Sir Gawain as a fictional Arthurian knight in *SG*), or else in the broad evaluative painting of a particular social group or profession, with friars and other clergy members popularly featuring in these types of negative consumption-based evaluation. This approach blurs the line between real-world and literary representations with evaluation operating as 'an ambiguous semantic space between demonstrable material practice and symbolic meaning' (Jones 2011: 16).

Based on these results, it would not be unreasonable to infer a more general divide in poetic approach to consumption as a broader topical theme. However, validation of this would require further support though more detailed analysis of the other consumption-based keywords, which lies outside the scope of this immediate study.

5.5 Conclusion

The present chapter provided a pattern-driven approach to the language in which themes of consumption are approached in the text. Keyword analysis identified the consumption-based words appearing with statistical significance in *PP* compared to the

other medieval texts within the reference corpus, while collocational analysis enabled detailed study of the underlying linguistic mechanisms contributing to specific effects in the text (the generation of figurative meaning in *hunger* and *glutton* and the evaluative processes linked to the negative semantic prosody associated with *drink*). This analytical focus on *hunger, glutton,* and *drink* as bridging or 'split' words in Mahlberg and McIntyre's (2011) KW framework enabled this chapter to be oriented around those consumption-based KWs demonstrating the most ambiguity in their textual functionality, and thus words likely to be of high stylistic potential in their use.

The analysis of *hunger* and *glutton* using Hoey's (2005) theory of lexical priming in 5.3 identified the animacy-based linguistic methods through which additional layers of pragmatic metaphorical significance are attributed to instances of personification allegory in the poem. These approaches were shown not to be unique to *PP* but are particularly useful in demonstrating the underlying mechanisms contributing to the presence of inferential ambiguity in certain contextual uses within medieval poetry in general.

Analysis of *drink* using the evaluative framework provided by Martin and White (2005) in 5.4 revealed the judgement-based valuations of social condemnation characteristing the poem, and highlighted the heavily negative discussions of morality arising from the contextual uses of this keyword. This prototypical representation of drink(ing) in *PP* provided a unique basis for which to start exploring comparisons between the similarities and differences evidenced in the reference corpus. The predominantly negative morality-based evaluation observed in *PP* was ultimately shown to be less characteristic of *MVCorp* than *ARCorp*, which suggests fundamentally different ways of approaching consumption in medieval poetry based on the influences on (and content ultimately discussed within each) poetic text-type.

6. LEXICAL REPRESENTATION OF THE LAW

This second analysis chapter contains an exploratory examination of the legal language represented in the poem's KWs. As highlighted in 2.1.6.2 *PP* contains unusually high amounts of legal terminology, which makes legality a theme of topical interest in the text. The current chapter builds on existing study of legal language in *PP* by demonstrating how corpus based frameworks and analyses can offer new insights into the linguistic patterns used to textually portray these legal concepts. This study does not intend to provide a systematic exploration of all the legal language in *PP*, but rather to investigate the stylistic impact of its comparatively frequent legal terminology, and to furthermore demonstrate how these effects are achieved on a textual level.

6.1 starts by introducing the legal terminology identified in the poem using Alford's (1988a) glossary of legal diction. 6.2 then organises these legal KWs into the functional categories offered by Mahlberg and McIntyre (2011), which reveals new features and categories not evidenced in the consumption-based lexis. Within this categorisation, the stylistic potential revealed in the poem's legal split-words is further investigated in 6.3 by focusing on the KWs *doctor* and *justice* as words demonstrating dual fictional and thematic functions in the poem as examples of class allegory. 6.4 rounds out the chapter by examining the stylistic potential in *reason*, whose multiple fictional world and thematic textual functions can be linked to its role as a sub-techical legal term along a cline from highly technical to more general terminology with occasional legal applications (Marin & Rea 2014, Marin 2014).

6.1 Keywords

From among the 855 keywords generated for *PPCorp*, 119 have been identified as belonging or relating to medieval legal discourse. These legal KWs were collected by comparing the *PPCorp* KW list to the entries in Alford's (1988a) legal glossary, with the resulting listing including terms which are used in a legal context in at least 50% of their textual concordances. To this end, Table 6.1 does not contain all of the legal terminology appearing in the poem, but rather features concepts shown to appear with unusual frequency compared to the given reference corpus.

The range of keyness values revealed in Table 6.1 illustrate the relative significance of these legal KWs to the poem. The entries appearing at the top of the table represent the most statistically significant legal terms to (and within) the poem, featuring concepts such as *conscience* (12) and *mede* 'meed' (17). These two keywords possess unusually high keyness values at 495.41 and 337.97 each, and head a small list of legal terms retaining keyness values well over 100, including recognisable legal terms such as *law* (40), *truth* (41)*, reason* (45)*, witnesses* (47)*,* and *mercy* (54).

Table 6.1: List of the 119 total legal KWs in *PPCorp*

N	**Key word**	**Keyness value**
12	*conscience*	495.41
17	*mede*	337.97
30	*ignorant*	194.26
40	*law*	156.75
41	*truth*	152.67
45	*reason*	143.00
47	*witnesses*	141.99
54	*mercy*	121.40
55	*clerks*	121.07
61	*dei*	106.76
66	*saved*	101.91
68	*cattle*	99.60
69	*pardon*	99.60
71	*pees*	98.64
76	*loyal*	87.05
82	*penances*	81.17
88	*writ*	77.36
102	*grace*	66.67
115	*leave*	59.90
116	*suffer*	59.87
120	*soffrede*	57.42
140	*cam*	50.82
142	*ad*	50.77
145	*symonye*	50.69
148	*learned*	49.27
150	*amend*	48.71
170	*lords*	44.77
175	*doctor*	43.66

193	*witness*	41.19
194	*privy*	41.17
195	*guilt*	41.17
197	*movable*	40.63
198	*possessions*	40.63
213	*help*	38.46
223	*syuyle*	36.54
227	*rect*	36.54
231	*nolite*	35.63
233	*god*	35.45
241	*perpetual*	33.94
243	*thieves*	33.87
244	*reave*	33.87
265	*matrimony*	32.54
269	*leely*	31.32
276	*curatours*	31.32
285	*faith*	31.11
289	*witiness*	30.66
297	*penance*	29.03
303	*justice*	27.76
305	*shows*	27.53
306	*true*	27.50
325	*akyng*	26.10
330	*iouste*	26.10
333	*pursued*	26.10
343	*euenesse*	26.10
345	*tithe*	26.10
353	*contreuede*	26.10
361	*soffred*	25.73
363	*conseyl*	25.73
368	*false*	25.35
374	*si*	24.56
378	*kings*	23.66
383	*comes*	23.53
390	*parsons*	22.63
391	*justices*	22.63
419	*bondsmen*	20.88
426	*release*	20.88
431	*curator*	20.88
432	*conseyle*	20.88
435	*restitucioun*	20.88
440	*sysores*	20.88
442	*syuile*	20.88
462	*iustus*	20.88
463	*hayward*	20.88
466	*indulgences*	20.88

476	*mayres*	20.88
482	*peccatum*	20.84
484	*legem*	20.84
494	*pena*	20.84
496	*claims*	20.80
514	*for*	19.36
515	*lordship*	19.30
529	*soeth*	18.03
544	*pays*	17.03
547	*bulls*	16.92
551	*me*	16.70
555	*save*	16.13
563	*notaries*	16.03
567	*arrears*	16.03
575	*payede*	16.03
579	*sekte*	16.03
582	*allowed*	16.03
584	*simonye*	16.03
588	*nomine*	15.97
593	*treuth*	15.97
595	*culpa*	15.97
598	*seal*	15.97
601	*alawe*	15.66
602	*commissary*	15.66
607	*claymen*	15.66
624	*bastards*	15.66
628	*borweth*	15.66
644	*apose*	16.66
646	*caes*	15.66
651	*vsure*	15.66
658	*pauper*	15.66
661	*lordeyne*	15.66
675	*fellowship*	15.66
678	*faught*	15.66
679	*faytour*	15.66
693	*y-maried*	15.66
702	*manibus*	15.66
703	*iudicij*	15.66
704	*forgiveness*	15.66
711	*debes*	15.66
713	*reddet*	15.66
716	*rihtwysnesse*	15.66
731	*raueschede*	15.66
752	*quasi*	15.66
756	*non*	15.45

The majority of the legal KWs for *PP* possess notably smaller keyness values than this top group, with over a third in Table 6.1 appearing towards the bottom of the keyword list and concentrated within a keyness range between 15 and 21. This less statistically significant group of keywords contains a number of terms relating to specific legal professions referenced in the poem such as *sysores* 'azzizers' (440) and *notaries* (563), as well as held offices such as the civil *mayres* 'mayors' (476) and the ecclesiastical *commissary* (602). These lower-keyness entries still represent central legal concepts (e.g. *caes, arrears, restitucioun*) and processes (*claymen, borweth*); however, these terms would also be expected to more commonly feature within other texts in the reference corpus alongside *PP*.

As with *PP*'s consumption-based key lexis, a number of legal KWs listed in Alford (1988a) were not included in the present study. In addition to exclusions made on the basis of concordance analysis revealing less than 50% of textual occurrences relating to legal content (*both, makes, courtesy, lord*), KWs such *absolved* (406) were removed from Table 6.1 based on diachronic changes in the legal significance attached to these word forms in a historical context. The legal meaning of the keyword *absolved* would seemingly correspond with Alford's head word ABSOLUCION. However, rather than including the verbal meaning 'to release from an oath or obligation' (*MED*), Alford (1988a: 1) only extends the legal meaning of *absolucion* to its capacity as an official legal document granting remission from sins committed, hence the exclusion of *absolve* and *absolved* from the finalised keyword list. Keywords in *PPCorp* similarly approximating but not representing medieval legal terminology include *leaves* (404), *pestilences* (279) and the various forms of *give* (85, 104); these do not appear in the list of legal keywords within the poem, nor are they addressed in the rest of the chapter.

6.2 KW classification by text function

To start examing the ways in which these terms are utilised stylistically in *PP*, the legal terminology identified in Table 6.1 was categorised into Mahlberg and McIntyre's (2011) functional categories. This step reveals the ways in which the law in *PP* is represented through its fictional world-building elements, as well as the range of thematic signals conveying legal significance beyond the identification of legal professionals, settings, and props within the text. As done with the consumption KWs in 5.2, categorisation was conducted through analysis of each KW's concordance lines to determine whether the majority of their uses in *PP* function as fictional world elements or thematic signals.

Figure 6.2 below reveals the key fictional world and thematic signals used to depict legality in *PP*. As seen with the consumption-based KWs in Figure 5.2, the majority of these terms constitute fictional world signals sub-divided into categories relating to 'characters', 'settings and props', 'laws', 'descriptors' and 'actions', with a comparatively smaller grouping of identified thematic signals.

Figure 6.2: Fictional world and thematic KWs in *PPCorp*

Category	Example key words
Fictional world	
Characters:	
Names	*conscience*, mede*, reason*, mercy*, faith*, false*, syuyle/syuile** 'civil', *grace**
Titles and legal positions	*doctor*, reave*, justice/iustus*, counseyl/conseyle* curatour*, hayward**
Descriptors	*ignorant* (lewed), *learned* (lered), *loyal* (lele), *leely*
Collective referents	*thieves, curatours, kings, parsons, justices, bondsmen, sysores, mayres, sekte, bastards, lords, fellowship, witnesses, clerks, notaries, counseyl/conseyle**
Settings and props:	
Documents and tools	*writ, indulgences, bulls, seal, pauper* 'paper'
Laws and legal maxims:	*law/alawe, dei, ad, help, nolite, god, si, peccatum, legem, pena, me, nomine, manibus, iudicij, debes, reddet, quasi, non, perpetual, syuyle/syuile**
Legislated activities and domains	*symonye/simonye, cattle, movable, possessions , tithe, arrears, vsure, lordship*
Actions:	*cam* (comen), *witness*, *comes*, *iouste, pursued, apose* 'inquire/investigate', *claims/claymen*, *allowed, borweth, faught, y-maried, reason*, raueshede* 'rob',

	contreude
Punishments	*penances, amend, penance, restitucioun*, pays/payede, save*(*d*)
Pardons	*mercy*, pardon, for(ȝeven)/euenesse* (*forȝeueness*), *release, forgiveness**
Thematic signals	
	conscience, mede*, reason*, mercy*, faith*, guilt, false*, treuth, doctor*, reave*, justice/iustus*, soeth, true, restitucioun*, counseyl/conseyle*, forgiveness*, hayward*, curatour*, rihtwysnesse*
Uncategorised	
	caes ‘case’, *privy*

*denotes words split between different functional categories

The first of the fictional world signalling categories, ‘characters’, is comprised of three sub-groupings of KWs. The first category contains a selection of common nouns and some adjectives operating as the names of the poem’s major (*conscience, mede, reason*) as well as minor allegorical characters (*faith, false*, and *syuyle* ‘civil’). Within this group, *mercy* is a unique term in that it operates as the referent for multiple characters across different parts of the text. The ‘titles’ sub-category contains singular common nouns denoting legally-recognised positions and offices including court-oriented positions (*justice*, *doctor*) in addition to the legal offices seen more locally on manorial estates (*reave*, ‘steward/overseer’, *hayward* ‘property officer’) (Alford 1988a: 67, 135). The ‘descriptors’ section also includes adjectives charactersising persons as

being *lewed* (unlearned), *lered* (learned), or *lele* (loyal or legal), which have set connotations within a legal context.

The 'collective referents' sub-category, as demonstrated with the consumption-based KWs in 5.2, is comprised of plural common nouns which identify a collective group of persons. These referents allude to persons holding legal power or office (*kings, justices*, *mayres* 'mayors', *sysores* 'azzizers', *counseyl* 'legal counsel', *clerks, notaries*), holding legally-recognised positions in the religious sector (*parsons, curatours*), operating under legal contract (*bondsmen*), enactors in the courtroom (*witnesses*), those recognised via legally established standards of legitimacy (*bastards*), or whose activities flout the law in some capacity (*thieves*). This category also contains three collective nouns referring to groups of individuals, with *fellowship* referring to a group of legal officers (1988a: 57), *sekte* a retinue or party of followers involved in litigation (1988a: 140), and *counseyl/conseyle* in the majority of its uses in *PP* referring to the group of legal experts comprising the king's legal council (1988a: 37-8).

By contrast, the 'settings and props' category in Figure 6.2 contains a comparatively small amount of the poem's legal KWs. There are virtually no indicators of physical setting (e.g., courts) but there are references to legal props in the poem's KWs, particularly through nouns relating to legal documents (*pauper* 'paper', *writ, indulgences*) and the tools denoting their authenticity (*seal, bulls* 'lead seal'). This absence of setting-based terms was also reflected in the 'setting and props' category back in Figure 5.2, which only features props in the form of consumable products.

The 'laws and legal maxims' category continues the fictional world development of *PP* through recourse to the law and the domains under its regulatory power. This category is a new addition not seen in the consumption KWs and contains most of the direct textual references to *law*, including the textual conflation of the

noun and indefinite article in *alawe* ‘a law’. The remainder of the KWs contained in this category consist of a combination of English and Latin lexis appearing in legal maxims such as such as *legem* in *necessitas non habet legem* ‘need has no law’ (Alford 1988a: 102), as well as formulaic English legal phrases such as the traditional witness’ oath *so me god help* (1988a: 146). The legal significance of these KWs are based on their phrasal use as much as their legal denotational meaning, as reflected below in Figure 6.3 with the textual use of the Latin term *pena* ‘penance’ when discussing culpability (1-3) and maxims of canon law (5) such as *pena pecuniaria non sufficit* [*pro spiritualibus delictus*] ‘pecuniary penance does not suffice for spiritual faults’ (1988a: 113).

Figure 6.3: Concordance of the five occurrences of *pena* in the annotated *PPCorp*

1	And pardon with the ploughman A	**pena**	& a culpa And all holy Eremytes
2	the margin had many years Ac no	**pena**	& a culpa no truth would hem gra
3	earth And purchased him a pardon	**pena**	& A culpa For him and his airs f
4	ing Consencientes & agentes pari	**pena**	punientur Patriarchs and prophet
5	or the pol by pol if you be rich	**Pena**	peccuniaria non sufficit &cetera

This law-oriented category also features a sub-category containing predominantly nominal KWs illustrating the activities and domains benefitting from said legal oversight. Terms such as *vsure* ‘usury’, *arrears, symonye* ‘simony’, *lordship, tithe,* and *possessions* reflect legal interest in overseeing financial transactions and property rights for both civil transactions and those within ecclesiastical contexts, with *cattle* ‘livestock’ and *movable* denoting types of legally-recognised property.

The ‘actions’ category represents the final main world-building category in the thematic representation of legality in *PP*. Predominantly comprised of verbs, this category provides an inside view of typical activities performed in the courts like *cam/comes, witness,* and *claims/claymen*, despite the lack of direct allusion to said

settings within the poem's KWs. They also refer to processes occurring beyond the courts under legal remit (e.g., *borweth* 'borrows' and *y-maried* 'to endow for marriage'), as well as actions with legal repercussions (*raueshede* 'rob'). From amongst these terms *iouste* 'joust' and *reason* stand out as isolated terms, with the former a method of metaphorically representing a judicial battle common in religious motifs (Alford 1988a: 76) and the latter operating as a type of cognitive action compared to the mostly physical or wordly actions otherwise characteristic of this category.

Within 'actions', the sub-categories listed in Figure 6.2 reveal the primary outcomes of these legal actions as being oriented around punishment and pardon as two methods of achieving a legally-recognised form of settlement. Key lexis involving punishment in *PP* tends to focus on financial *restitucioun* – actions intended to restore ill-gotten property or making more general amends for a sin, crime, or debt (Alford 1988a: 135) – through actions involving recompense (*pays/payede*) either by the charged individual or by another on his behalf (*saves*) (Alford 1988a: 107, 139) as a process of making amends (or, in cases where additional penalties are warranted, conducting *penance*). *Pardon*, by contrast, is mostly centred on the nominal and verbal manifestations of *mercy* (1988a: 99) and *foregiveness* [*foryeven*] (1988a: 63) as methods of clemency (1988a: 109) and thus a person's *release* from legal obligation (1988a: 131). *Mercy* operates as a highly key term in *PP* and *foregiveness*, while a key concept to (and in) the text, is less clearly represented in the corpus data due to the range of split word forms (e.g. *for* + *ȝeven/euenesse*) resulting from yogh-based recognition issues in *WordSmith*.

Compared to the fictional world elements, the thematic signals are represented by fewer overall key terms in *PP*. This category features abstract nouns which, in addition to being examples of property allegory (e.g., *conscience, mercy, false, grace,*

mede) also represent core aspects of law that have wider thematic import including *rihtwysnesse* 'rightfulness', *true*, and *soeth* 'truth'. *Forgiveness* and *restitution* are terms which similarly, in a text like *PP*, also ostensibly possess additional thematic significance beyond their role as real-world legal processes. While in the vast minority compared to the poem's demonstrated world-building lexis, these thematic signals do themselves outnumber some of the smaller fictional world categories like 'settings' and character-based 'descriptors'.

As with the consumption-based lexis, the stylistic potential exhibited in these functional KW groupings is oriented around terms split across these fictional world and thematic categories. Figure 6.2 reveals a higher number of split-words (12) compared to the consumption KWs (5), which is likely indicative of the higher numbers of legal KWs generated for the text (119 versus the 26 in Chapter 5). Property allegory, as well as terms such as *doctor, reave, hayward*, and *justice,* lie on the boundaries between the world-building 'characters' and thematic signalling categories, with *counseyl/conseyle* additionally operating in multiple fictional world categories based on its use as both a common and collective noun referring to either a court advocate or a member of the king's legal counsel, respectively (Alford 1988a: 37-8).[61] *Mercy* and *reason* notably stand out as terms exemplifying unusual functional dynamicity in their use as split-words, with both terms operating not as only abstract concepts in a thematic sense (mercy as ideological property) but also as an allegorical referent (Mercy) and an action (e.g., being merciful or bestowing mercy on someone).

The remainder of the chapter focuses on three of these split-category KWs: *doctor* and *justice* as examples of terms possessing dual fictional world and thematic

[61] Civil (*syuyle, syuile*) is another KW bridging multiple fictional world and thematic categories through its use to describe a type of legal system (e.g., *civil law*), in addition to being an example of property allegory.

functions (6.3), and *reason* as an example demonstrating high amounts of multi-functionalism in the text (6.4).

6.3 Dual functionality via class allegory: *doctor* and *justice*

Like *hunger* and *glutton*, the KWs *doctor* and *justice* possess both fictional world and thematic text functions due to being types of personification allegory. Instead of appearing as proper nouns denoting the name of a fictional character, however, *doctor* and *justice* are common nouns that identify a specific but otherwise unnamed character. To this end these KWs operate as examples of class allegory, a sub-category of personification allegory which Crisp (2001: 14) defines as the personification of a demographic or socially-constructed group of persons as a single entity through grammatically singular common nouns.

While similar to collective nouns like *clergy* and *community*, class allegory is different in that it directly identifies an unnamed character via a characteristic feature ascribed to the person instead of through a given name. In the case of *doctor* and *justice*, these allegorical constructs demarcate persons via their professional roles as legal officers, with *doctor* being a lawyer of canon law (Alford 1988a: 48) and *justice* referring to a judge or other legal officer of the court (1988a: 76). This approach of identifying unnamed characters by legal, or legally recognised, occupation is also observable in Figure 6.2 in KWs such as *reave* 'reeve', *hayward* 'overseeer' and select uses of *counseyl/conseyle* 'counsel'; it also more clearly classifies the previously non-categorised consumption KW *brewer,* as listed back in Figure 5.2.

To explore how the text's class allegory contributes to the stylistic representation of legality, 6.3.1 first identifies the main collocates for *doctor* and

justice in *PPCorp*. 6.3.2 follows by ascertaining how these fictional world and thematic signals are linguistically encoded in *PP* through the lexical priming achieved via their collocates. 6.3.3 links this content to the surrounding literary criticism via work conducted on theoretical frameworks modelling allegory, while 3.3.4 concludes by comparing the collocational patterns evidenced in *PP* to their textual uses in *MEPoetLitCorp*. Throughout the course of this study (6.3.1-6.3.4), I build on the functional groundwork for allegorical functionality created in Chapter 5, with alterations and additions to my original classifications clearly signposted where needed to better take into account this new type of allegorical personification not present in the poem's consumption-based KWs.

6.3.1 Collocations

In this section I first identify the main lexical and grammatical collocates for *justice* and *doctor* in *PPCorp*. These collocates were generated through the concordance tool in *WordSmith* using mutual information (MI) with a L3-R3 window of proximity. The aim in generating these collocates is to show these KWs to be part of larger textual patterns in the poem, which are then linked back to the different text functions (fictional world and thematic signals) in the next section (6.3.2). The discussion contained here focuses on the textual use of *doctor* and *justice*, with both the observed patterns and import of said patterns more broadly representative of the trends observed for the other types of class allegory present in the text's KWs.

As evidenced below in Tables 6.4 and 6.5, the main lexical collocates for *justice* and *doctor* are primarily comprised of verbs and nouns. *Doctor* collocates most frequently with *said*, followed by verbs such as *sethe* 'say', *oppose* 'ask'*, suffered, died,*

know, drink, preached, and *conieyed* 'got rid of'. This verbal collocation is also present with *justice* through textually proximal terms such as *told, declare, jogged, jug* 'judge', *ransomed, save, assent*, and *iustede* 'jousted'. The nominal collocations for these KWs are divided into proper nouns demarcating biblical (*ioseph* 'Joseph', *jesus*), historical (*dido*), and *PP*-specific fictional figures (*dowel* 'Do-well', *dobest* 'Do-best') in addition to a single place-name (*egipte* 'Egypt'). These collocates occur alongside a wealth of abstract nouns including *worth, power, death, dearth, truth* and *love* and a further selection of concrete nouns including *jews*, *barnbuyrne* 'man', *men, sire, tale, days* and *dyuynour* 'diviner'. *Kings* is a particularly interesting collocate in that it does not operate as a plural common noun (i.e., several kings) but rather as an inflected genitive form denoting possession (or, in the case of *kings justice*, a type of secular law overseen by the king and his council).

Tables 6.4 and 6.5: The 45 and 51 collocates for *justice* and *doctor* respectively in *PPCorp,* organised by textual position relative to the node word and listed in order of collocational affinity by raw frequency.

LEFT lexical		RIGHT lexical
kings (2), *jews* (2), *chief* (1), *truth* (1), *told* (1), *barnbuyrne* 'man' (1), *declare* (1), *dearth* (1), *ioseph* (1), *o* 'one' (1) *jogged* (1)	*justice*	*jesus* (2), *jug* 'judge' (1), *power* (1), *ransomed* (1), *said* (1), *save* (1), *worth* (1), *false* (1), *egipte* 'Egypt' (1), *assent* (1), *dawn* (1), *latin* (1), *men* (1), *iustede* 'jousted' (1)
LEFT grammatical		RIGHT grammatical
a (5), *be* (4), *before* (2), *until* (2), *was* (1), *and* (1), *of* (1), *that* (1), *nor* (1), *the* (1), *over* (1), *or* (1),	*justice*	*and* (4), *that* (2), *for* (2), *to* (2), *y* 'I' (1), *be* (1), *the* (1), *was* (1), *how* (1), *have* (1), *as* (1), *in* (1), *if* (1)

LEFT lexical		RIGHT lexical
said (3), *dowel* (2), *dobest* (2), *days* (1), *sire* (2), *oppose* 'ask' (1), *suffered* (1), *dido* (1), *died* (1), *conieyed* (1)	*doctor*	*said* (2), *dowel* (1), *days* (1), *sethe* (1), *preached* (1), *red* (1), *tale* (1), *death* (1), *love* (1), *three* (1), *know* (1), *drunk* (1), *drink* (1), *dysores* (1), *dyuynour* 'diviner' (1)
LEFT grammatical		RIGHT grammatical
this (7), *that* (2), *and* (1), *is* (1), *a* (1), *of* (1), *as* (1), *for* (2), *to* (1), *yourself* (1), *there* (1), *thus* (1), *do* (1), *art* (1), *himself* (1), *here* (1), *soon* (1), *son* 'soon' (1), *ever* (1)	*doctor*	*and* (4), *is* (3), *he* (3), *that* (2), *a* (2), *of* (2), *as* (2), *y* 'I' (1), *the* (1), *so* (1), *dere* (1), *are* (1), *at* (1), *be* (1), *hit* (1)

Each KW is also shown to collocate with several adjectival forms. The collocating adjectives of number (*o* 'one', *three*) and *false* do not end up textually characterising either KW (e.g., *three days*, *o barnbuyrne* 'one man', and *false latin*) but *chief* is used to modify *justice*, while *red* is used as a metaphorical comparative benchmark to illustrate the flush in the cheeks of a particular *doctor* in the text, as illustrated in (14) in Figure 6.7 below.

In addition to these lexical collocates, *doctor* and *justice* are shown to collocate with a selection of closed-class grammatical words in Tables 6.4 and 6.5. As with the consumption-based *hunger* and *glutton*, these KWs recurringly find themselves appearing in textual proximity with conjunctions (*and, or, nor, how* 'how that'), personal pronouns (e.g., *he, y* 'I', *hit* 'it'), prepositions (*for, of, to, at, before, over, until* 'until/towards'), demonstratives (*this, that*), modals (*have, do, are/is/art/be*), definite and indefinite articles (*the, a*) and some adverbs (*soon, thus, ever*). *Doctor* and *justice* also evidence additional collocation with reflexive pronouns (*himself, yourself*), features which were not present in the collocations for the property allegory studied

earlier in this investigation. The collocational patterns for each KW are better visualised in the concordance lines listed below, with Figure 6.6 revealing the collocational patterns for *justice* and Figure 6.7 the patterns evidenced in the use of *doctor* in the text.

Figure 6.6: Concordance of the 11 total occurrences of *justice* in *PPCorp* generated in *WordSmith Tools*, ordered by L1 position

N	Concordance
1	d wisdom of Westmunstre hall He jogged until a justice and iustede in his ere And ouertulde al his t
2	um selde sit poverty the truth to declare Or a justice to jug men me enjoins thereto no pore Nor to
3	on And then shall death withdraw and dearth be Justice And dawn the dividing day for sin But if god
4	Lack hem not but let hem worth until truth be Justice And have power to pynschen hem then pot forth
5	Al shall be but a covered and o barnbuyrne be Justice And that worth true tongue atydy man that ten
6	ding A Charter is challengeable before a chief Justice if false latin be in that letter the law hit
7	or king of the kyngdoem of Iuda Nor over Jews Justice as jesus was hem thought Here-of had Cayphas
8	itness Any creature be coupable before a kings justice Be ransomed for his repentance there all reas
9	ent And Conscience in all my Courts be a kings Justice y assent said reason by so yourself here Audi
10	ree days after edify hit new The jews told the justice how that jesus said Ac the overturning of the
11	s fadur said in farao his time That Ioseph was Justice Egipte to save his eleuene bretherne him for

Figure 6.7: Concordance of the 14 total occurrences of *glutton* in *PPCorp* generated in *WordSmith Tools,* ordered by L1 position

N	Concordance
1	e merry with this mete Ac y mourned ever For a doctor that at the hey deys drank wine fast Ve vobis
2	t you has brow now brouk hit yourself That art doctor of death drink that you made For y that am lor
3	and comsynges of dowel Of dobet of dobest for doctor he is know And over skripture the skilfole and
4	if the lik dowel For of dobet and dobest here doctor is dere love Tho went y my way without more le
5	ity and toward me he looked What is dowel sire doctor said y is dobest any penance dowel said this d
6	ence to pray me be still And said himself sire doctor be so hit be ?our will what is dowel and dobet
7	hat he full in a paralysis And there died that doctor are three days after Now y see said life that
8	science y took keep conieyod conieyed son this doctor And sethe a said to clergy so that y hit heard
9	d love as you techest This is a dido said this doctor a dysores tale Al this wit of this world and w
10	f and male-ese christ for man suffered As this doctor and dyuynour said y and decretistre of Canon A
11	tience hit heard Hit is not three days do this doctor that he preached At paul's by-fore the people
12	er And then is time to take and to oppose this doctor Of dowel and of dobet and if dobet do any pena
13	r said y is dobest any penance dowel said this doctor and he drunk after do thy neighbor non harm no
14	sat still as patience would and thus soon this doctor As red as a rose reddened his cheeks Coughed a

In looking at Figures 6.6 and 6.7, one of the most striking collocational patterns involves the use of articles to signpost definiteness and indefiniteness in the contruction of these concepts in the poem. This is a particularly prominent pattern in the L1 position (the term immediately preceding the node word), with 9 out of the total 25 occurrences of these KWs involving grammatical definiteness in the co-occurrence of *that* and *this* with *doctor* (7-14 in Figure 6.7) and *the* with *justice* (10 in Figure 6.6). *Justice* evidences higher comparative use of indefinite constructs with the article *a* in five textual occurrences as opposed to the single use in the phrase *a doctor* (1), though three of the five appear in the L2 position within the syntactic construction *a* + ADJECTIVE/POSSESSOR + *justice* (6, 8, and 9). This high recurrence of patterns based on definiteness stands in contrast to the recurring binomial patterns and verbal colligations for *hunger* and *glutton*, though there is evidence of grammatical subjecthood in the collocation of *doctor* and verbs like *said* in (9) and (13). The presence of definiteness as a new collocational feature suggests that definiteness may have a more contributory role in the lexical priming of the poem's class allegory than its property allegory, a concept explored in greater detail below in 6.3.2.

6.3.2 Lexical priming

As discussed in 5.3.2, lexical priming offers a method of linking a term's collocational patterns to the types of inferential meaning primed in the interpretation of these allegorical keywords. Within the poem's property allegory verbal colligations signposting grammatical subjecthood were shown to prime interpretations of *hunger* and *glutton* as world-building character referents, while L1 collocation with possessives primed understandings of the terms as ideological concepts of thematic

significance to the text. In both patterns, the presence or absence of agency was localised as the determining factor in which functionality was textually encoded in each use, with animacy offering a more nuanced approach to the levels of clarity and foregrounding involved in distinguishing these text functions from one another in instances of ambiguity or passages characterised by unusual stylistic creativity.

The poem's class allegory, as represented through the KWs *doctor* and *justice*, follows property allegory to some degree in having their pragmatic meaning based on perceived agency. *Doctor* appears frequently within verbal constructs demarcating the KW as the grammatical subject in *this doctor that he preached* 'this doctor preached' (11), as well as the object in the phrase *oppose this doctor* 'ask this doctor' (12). This colligation does involve select instances of physical activity with figures that drink wine too fast (1) and ones that even die (7), but most of these patterns represent more cognitive actions like speaking (5, 6, 9, 13) asking (11), and preaching (12).

However, agency-via-collocation is less significant a feature in class allegory than in property allegory. This difference in the importance of conveyed agency in the priming of class allegory is based on the disparate levels of semantic animacy embedded in the allegorical referents themselves, with animacy representing the cognitive distinction between the animate and inanimate operating along a hierarchical gradient (Yamamoto, 1999: 149) distinguishing between human entites and non-human concepts in allegorical constructs. Class allegory like *doctor* and *justice* are already encoded with some animacy in being an ill-defined human entity expressed by a common noun, regardless of the textual patterns within which they are situated; the description 'ill-defined' here suggests that better definition of a person can be achieved via use of a proper noun as an identified name. By contrast, properties such as *hunger* and *glutton* contain virtually no animacy embedded in the actual referents

being used in the representation of abstract concepts via inanimate nouns.[62] To this end, property allegory is more heavily reliant on the additional agency-based animacy provided by patterns of verbal colligation to achieve the more significant animacy-based shift from inanimate property to well-defined persons identified by proper nouns. This absence of need for verbally-conveyed agency would contribute to explanations behind the comparative absence of such actions as speaking and asking being attributed to *justice* observed in its concordances within *PP*.

6.3.2.1 Definiteness

Instead of the agency required for property allegory, the priming of class allegory is fundamentally more dependent on expressions of definiteness. Definiteness occurs in almost all textual examples of *justice* and *doctor* in *PP* with L1 collocation to (in)definite articles and demonstrative pronouns. Figure 6.7 reveals a high amount of L1 collocational patterns with demonstratives, with 8 of its 14 total occurrences in *PP* appearing in constructs such as *that doctor* (7) and *this doctor* (8-14). By contrast, Figure 6.6 shows *justice* to be more frequently found in indefinite collocational constructs in 5 out of its 11 textual uses. The placement of this indefinite article is shown to occur either the L1 position as with *a justice* (1-2), or in the L2 pattern *a* + DESCRIPTOR + *justice* (6, 8, 9). The balance of definite and indefinite patterns, if not evidenced at the KW level in this instance, is more clearly visible in a broader examination of the comparative percentages found in the other instances of class

[62] See Foley and van Valey's (1985) linguistic animacy scale in Figure 5.8 for a visual illustration of the different levels of animacy contained in inanimate nouns versus human common nouns.

allegory in the poem, resulting in approximately equal comparative collocation with definite and indefinite grammatical premodifiers.

How this (in)definiteness primes fictional world or thematic text functions in the poem's class allegory is illustrated below with *doctor*. The following textual passages contain instances where the KW collocates with function words conveying definiteness (i) and indefiniteness (ii), as situated within their greater contexts in the poem:

(i) And thus sone **this doctour**
As rody as a rose rodded his chekes
Cowhede and capede and consience hym herde
And tolde hym of a trinite and toward me he lokede (14)

*And thus soon **this doctor***
As red as a rose reddened his cheeks
Coughed and cleared his throat, and Conscience heard him
And told him of the trinity, and looked towards me… (15:113-6)

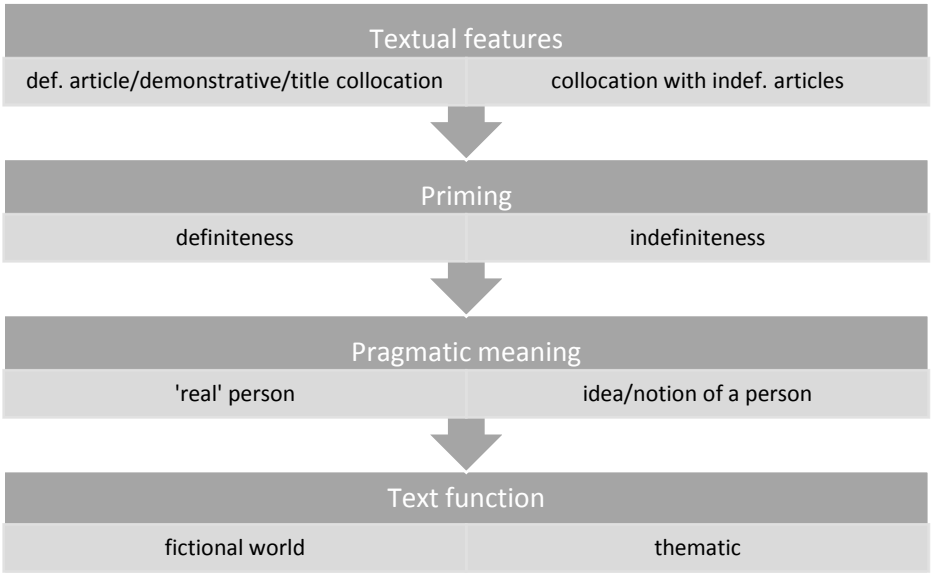
(ii) Ac y mournede euere
For **a doctour** þat at þe hey deys dranke wyn faste
…And ete manye sondry metes mortrewes and poddynges (1)

But I mourned without end
*Because **a doctor** at the high dais drank wine fast*
…And ate many sundry meats and puddings (15:67-9, 70)

The textual definiteness generated from *this doctor* in (i) primes an allegorical understanding of *doctor* not achieved in the indefinite construction *a doctor* in (ii). Definiteness cognitively orients focus on a 'real' (albeit, still poorly-identified) person that stands apart from the range of other potential doctors that could ostensibly exist within the late Middle English landscape. This pragmatic differentiation between a specific person versus the conceptual *notion* of a person, based on the presence of textually-encoded definiteness, foregrounds the world-building functionality of *doctor*

as a real, but ultimately unknown or not well-known fictional character in (and to) the text; in other words, a minor character in the poem.

Figure 6.8: Update to Figure 5.7 illustrating how definiteness primes textual functionality in class allegory, moving towards increasing levels of abstraction.



In addition to definiteness, fictional world functions are also shown to be foregrounded in instances involving collocation with personal titles. The recurring L1 collocation of *sire doctor* (5, 6), for example, achieves pragmatic realisation of a specific character in the text, albeit with a better-defined identity attributed to said figure than generated for the construct *this doctor*. Referential titles like *sire* and *madam* do not fit neatly into Foley and van Valin's (1985) linguistic animacy scale but rather operate as features situated in an intermediary stage between human common nouns and the more animate personal pronouns and proper nouns category. In this sense, the additional animacy attributed to the perceived figure of *sire doctor* is based not on higher degrees of agency but rather closer degrees of relational affinity and specificity. The agency associated with *sire doctor* can be increased with the additional verbal colligation denoting the doctor as the person being addressed as a type of

personhood seen in both (5) and (6), but the definiteness in these cases is itself sufficient to generate a world-building text function as illustrated below in (iii):

(iii) What is dowel **sire doctour** quod y is dobest eny penaunce
Dowel quod **this doctour** and he dronke aftur
Do thy neyhebore non harm ne thy-sulue nother… (5, 13)

'What is Do-Well, sir doctor?' I said. 'Is Do-Best any penance?'
'Do-Well' said the doctor, then took a drink.
'Do your neighbour no harm nor yourself either…' (15:117-9)

Fictional world functions in class allegory can thus be said to partially rely on the strategic employment of different types of social deixis. The reference to social characteristics uniting or distinguishing between interlocutors (Levinson 1983: 63), social deixis is manifested through grammatical relationships like those evidenced in the L1 collocation of *this* and *sire* to *doctor* in *PP*. *This* operates as a form of relational deixis signposting the degree of relationship between referents, while *sire* exists as an absolute form of relational deixis wherein a social characteristic of the referent is used to set a referent apart through a relative ranking system commonly found in forms of address (Levinson 1979: 207). Distinguishing between these two textual methods of producing fictional world functions is useful in that it enables questions to be raised regarding the potential differentiation between minor characters that are relatively unknown but observed (e.g., *this doctor*) versus those potentially known but merely unnamed in the given context (*sire doctor*).

By contrast, to interpret the use of *justice* or *doctor* as a referent to a character rather than a thematic signal in the text's indefinite collocational patterns would be fundamentally at odds with its underlying lexical priming. As shown above in Figure 6.8, L1 collocation with the indefinite article *a* textually primes the indefiniteness leading to the pragmatic ideation of a hypothetical type of person representing a

certain profession. This representation of a notional person acting as a figural stereotype for the rest of a – in this case, professional – demographic foregrounds the thematic significance of its use, operating less as a description of a real (but unknown person) as it would do as a fictional-world element and more as a method of providing commentary on doctors as a profession. In other words, the underlying thematic information surrounding doctors and their evaluative significance in the late 14th century operating as a background conceptualisation in definite collocations is foregrounded in the referent's indefinite constructs.

This thematic function becomes clearer when viewing indefinite collocation as a textual method of enabling evaluative or otherwise critical commentary to be conducted without directly naming the individuals being targeted. This type of evaluation typically involves imbuing these notional allegorical figures with the values and habits stereotypically or ideationally associated with the given profession. This step is not unlike the attributes associated with the poem's property allegory like Truth and Reason; however, the characteristics associated with class allegory like *doctor* are more ideological than universal properties, with their representation dependent on the socio-culutral perception of the persons within these professions as positioned within a historical context in late medieval England. The targets of said evaluation could be the members of the given allegorised profession, as can be extrapolated in the description of the doctor engaging in overconsumption in (ii). This process can also be seen in the use of the position as a benchmark to which others do not meet, as exemplified in the cool appraisal of the merits of poverty in (iv):

(iv) …selde syt pouerte þe soth to declare
Or **a iustice** to iuge men me enioyneth þerto no pore
Ne to be amair ouer men ne mynistre vnder kynges
Selde is eny pore y-pot to penesche eny peple (2)

Seldom does poverty sit to declare the truth
Or to be a justice to judge men, no poor man is hereto enjoined
Nor to be a mayor over men nor minister under kings
Seldom is any poor man positioned to punish any people….
(16:126-131)

This evaluative potential aligns with Mahlberg and McIntyre's (2011: 209) expectation that metaphorical constructs, as well as those possessing evaluative import, can be generally classified as thematic signals in literary text.

6.3.3 Allegory in the literary criticism

The results of the present study of *doctor* and *justice* further build on the theoretical frameworks for studying allegory initially developed in the study of *hunger* and *glutton* back in Chapter 5. The current focus on a different type of personification allegory (class versus property) revealed new collocational patterns contributing to the textual use of allegory in *PP* as either fictional world elements or thematic signals, particularly though L1 collocation with deixis conveying (in)definite and socio-relational information. To reflect these new findings, Figure 6.9 below updates the original cline of interpretative animacy formulated in Figure 5.8, with (in)definiteness identified as a pragmatic marker of interpretative animacy alongside the agency and possession previously linked to (in)animacy in property allegory .

Figure 6.9: Hierarchy of linguistic, pragmatic, and cognitive animacy, as reproduced and further updated from Figure 5.8.

Cognitive animacy scale	***Langacker (1991)***	Speaker> hearer	Human	Animal	Physical object	Abstract Entity
	Stockwell (2009)	Specific persons	Ill-defined individuals >groups	Animals >plants	Machines> immovable objects	Landscape objects> abstractions
Interpretative animacy scale		Agency (movement>speech>sensory> sentience); definiteness			Absence of agency (possession); indefiniteness	
Linguistic animacy scale	***Foley & Van Valin (1985)***	1st & 2nd person>3rd person pronouns> proper nouns	Human common nouns	Other animate nouns	Inanimate nouns	

Interestingly, the allegorical meaning attributed to class allegory like *doctor* and *justice* stems from the opposite text-functions of of property allegory. The allegorical significance of terms like *hunger* lie in their fictional world functions as persons rather than inanimate existential states (or, in some cases, ideological properties). By contrast, terms like *doctor* are allegorical in their use as thematic signals inviting speculation about a group of persons operating within a set profession, as evaluated through the actions and behaviours ascribed to a stereotypically representational individual entity.

Figure 6.9 above thus sets out a model for studying personification allegory from a linguistic perspective. It provides the means by which to enrich studies of linguistic allegory in medieval literary studies by distinguishing form from function, and in doing so offers multi-dimensional analysis of personification and class allegory to assist in achieving greater overall analytical precision. It is worth observing that this cline cannot be said to represent all the language patterns through which allegorical

functions are realised; to make this claim, more study would be needed addressing other types of personification allegory (i.e, architectural), as well as examples gathered from other texts beyond *PP*.

Despite its currently limited applicability to certain types of allegory, the content of this theoretical framework has import to existing medieval scholarship examining allegory, both Langlandian and otherwise. It firstly corroborates research linking animacy to grammatical signals in allegorical personification signposted in studies such as Bloomfield (1970) and Griffiths (1985), and also aligns with the construction of other gradients including but not limited to Croft's (2003: 128-9) extended animacy hierarchy. This type of tiered animacy scale also offers an alternative paradigm for medieval scholars critical of rigid personification-based allegorical hierarchies based on grammar alone (e.g., Scanlon 2007), with the current approach presenting an aligned framework grounded in grammatical, pragmatic, functional, and cognitive interpretation.

In addition to corroborating and offering new approaches to allegory, the interpretative animacy scale in Figure 6.9 can also be used to further build on existing research in medieval studies. For example, Paxson (1994: 42-3) sets out a working taxonomy of personification based on what he refers to as 'figural translations', or types of textual personification. This classification system provides ontological categories separating the different processes of achieving personification, as documented below in Table 6.10.

Table 6.10: Outline of Paxson's (1994) taxonomy of figural translation

Category	Process
substantialisation, materialisation	Non-corporeal entity → physical one
anthropomorphism	Non-human → human
prosopopeia	Non-human → sentient human capable of thought and language
animification	Human → animal Or Abstraction → animal
ideation, rarefication	Human → abstraction
topification	Abstraction → geographical locus

Looking at the Paxson's (1994) categories, it can be observed that these categories are fundamentally oriented around – and thus, separated by – increasing and decreasing amounts of animacy. The first three categories (*substantialisation, anthropomorphism, and prosopopeia*) involve the progressive layering of increasingly stronger textual markers of animacy, while the middle two categories (*animification, ideation*) signposting animacy reduction. The final category, *topification*, is interesting in that animacy is unlikely to change in the figural translation from abstraction to geographic location but it will experience substantively higher amounts of definiteness in its grammatical transition from inanimate noun to place name. To this end, the hierarchical animacy scale revealed in the current investigation not only sets out the underlying ideological basis for Paxson's (1994) taxonomy but also provides the means to linguistically operationalise it by setting out some of the common lexical and grammatical markers of textual (in)animacy.

The following section (6.3.4) concludes the present discussion of *doctor* and *justice* by looking at the use of these KWs in the study's reference corpus to determine if the identified patterns of use found in *PP* are stylistic, or correspond to more broadly recognisable patterns indicative of the text's poetic register.

6.3.4 PPCorp vs. MEPoetLitCorp

Unlike the notable differences in how common the world-building function for *hunger* and *glutton* occurred between *PPCorp* and *MEPoetLitCorp*, *doctor* and *justice* generally appear to be used in a similar fashion across both literary corpora. For example, the 9 total occurrences of *doctor* in the reference corpus mirror the term's use in *PP* by existing solely within (in)definite collocational patterns. These patterns predominantly consist of L1 collocation (see Figure 6.11 below) and two conveying this definiteness in L2 patterns involving adjectival descriptors to futher characterise the given figures, with *holy* (6) suggesting a legally-recognised ecclesiastical official and noble (7) denoting the figure's high rank or otherwise laudable honour or moral character (*MED*).

Figure 6.11: Concordance of the 9 total occurrences of *doctor* in *MEPoetLitCorp*, as generated in *WordSmith Tools* and ordered by L1 position.

N	Concordance	File
1	that all samen on a sopp semblis the pupil, A doctor, one Domystyne they dearly beseke To consail	WA.txt
2	al sampnes in a sopp assempbles the people, A Doctor, one Domestyn they derly seek To Counsel the	WA.txt
3	cleped was the Maudelayne. With us there was a DOCTOR OF PHISYK, In al this world nor was there noon	GP.txt
4	er, He sat as a souvrayn on a high siege. A doctor of doutz, by dere God he semyd, For he had loo	MS.txt
5	as A clerk with a crown schauyn, And di?t as a Doctor in drabland fury, ?it all the earth of Egipt	WA.txt
6	can not bulte it to the bren, As can the holy doctor Augustyn, Or Boece, or the bishop Bradwardyn,	NPT.txt
7	his preface list to say; Solempnely this noble doctor dere Commends it, and says in this manner: The	SNT.txt
8	he calls, And as him dremyd same a dole that doctor he tells, And fraynes him fast on this fare ho	WA.txt
9	And glad is of his sorrow and his unhele; (The doctor makes this description). This maid up-on a day	PhyT.txt

Justice demonstrates similarity in its textual use between *PPCorp* and *MEPoetLitCorp*, but with some notable exceptions. Out of the 51 total occurrences of the term in the reference corpus texts, 6 uses appear in (in)definite constructions alongside the demonstratives *this* (1) and *that* (1) and the articles *a* (2) and *the* (2); *PP* is unique amongst these texts to ascribe definiteness to the term via terms of address (e.g., *sire*). The majority of uses involve *justice* as the quality associated with being fair or otherwise legally right (*MED*), uses which are textually manifested in patterns such as *with justice* (3), *of justice* (8), *if justice* (1), *in justice* (1), and of *his* (typically, God's) *justice* (4). It is briefly worth observing that, out of all of these uses of *justice* as a conceptual property, none of them appear within collocational patterns conveying the additional agency needed for interpretation as instances of personification allegory.

One pattern which does stand out is the appearance of *justice* as a verbal action. This concept of *justice* as an action stems from the Middle English verb *justisen*, which refers to the literal process of summoning someone to court for trial, or more broadly as types of righteous governance (*MED*). While none of the word forms for *justice* in the reference corpus are variants of this verb, the noun *justice* gets similarly situated within verbal phrases with some regularity; 9 out of the 51 total occurences appear within the construct *(to) do justice* or *done justice,* with 8 of these occurring in *Confessio Amantis* (*CA*) and one in *Romance of the Rose* (*RotR*):

(v) Kinghode may not well proceed
To do justice upon the right:
For it belongs to a knight… (12, *CA*)

However, despite the additional collocationally-conveyed agency, this use of *justice* is still fundamentally non-allegorical, not having the definiteness with which to signal the creation of a notional human entity.

To this end, the use of *justice* is overwhelmingly thematic in *MEPoetLitCorp*, despite the unusual presence of verbal colligation in its concordances. This to some degree mirrors the split thematic and fictional world functionalities present in the *PPCorp* concordances, through the presence of additional thematic patterns including L1 collocation with possessives (e.g., *his justice*) and the verbal *to do justice* indicate the higher amounts of variation present in the textual use of the term as a thematic signal. This departs from the use of *doctor* which, as signposted earlier, appears to be consistent in not only how it signals text function in *MEPoetLitCorp* through definite (e.g., fictional world) and indefinite (thematic) collocational patterns but also in the parity with which which each text function tends to appear in both corpora.

The second half of this chapter (6.4) continues the focus on split-words by shifting examination to the use of *reason* as a legal term possessing a thematic signalling text function, as well as two different types of world-building functionalities in the poem.

6.4 Multi-functionality in world-building and thematic signalling: *reason*

The third legal KW addressed in the current chapter is *reason*. As observed in 6.2 *reason* is used in *PP* as both a proper noun identifying the fictional character Reason, as well as a verb conveying the cognitive action of reasoning. The split-word *drink* examined in Chapter 5 shared this dual split fictional-world functionality, with *reason* also possessing an additional thematic signalling function as the ideological property of reason which, in a highly politicised text like *Piers Plowman*, ostensibly exists as a concept having greater overall significance to the poem than solely within the remit of its identified world-building functions.

The present section sets out to explore the different text patterns underlying the ways in which *reason* contributes to the development of the fictional word in *PP* as both a person and a performable action, as well as how it assists in developing thematic legality in the text. 6.4.1 examines the collocates generated for *reason* in *PPCorp*, which reveals the prototypical patterns within which each of these text functions occur and the high levels of technicality embedded in these collocations. 6.4.2 examines how this collocate-based technicality reveals broader patterns of use in the poem's legal language, with particular emphasis on the semantic creativity and polysemous potential present in semi-technical legal using Marin's (2014) classification framework. 6.4.3 contextualises the research conducted in 6.4.1-6.4.2 by comparing the concordances of *reason* in *PPCorp* to those generated for *MEPoetLitCorp* to determine differences in technical functionality, with the visual assistance provided through the use of lexical constellations (Marin & Rea 2014).

6.4.1 Collocations

The top grammatical and lexical collocates for *reason*, as generated from the annotated version of *PPCorp*, are reproduced below in Table 6.12. As with the approach taken with the other split KWs, these collocates were generated using MI within a 3-word span on either side of the KW in at least one occurrence. These collocates are listed below in order of their measured affinity to *reason,* with raw frequencies denoted in parentheses.

Table 6.12: Top lexical and grammatical collocates of *reason* in *PPCorp*, organised by textual position relative to the node word.

LEFT lexical		RIGHT lexical
right (13), *said* (15), *rich* (3), *truth* (3), *conscience* (6), *read* (4), *king* (4), *ride* (3), *kind* (4), *sire* (4), *rule* (1), *taught* (1), *rightful* (1), *riches* (2)	*reason*	*right* (12), *said* (5), *rich* (5), *truth* (5), *conscience* (2), *read* (3), *king* (2), *ride* (2), *pity* (5), *kind* (1), *rule* (3), *taught* (2), *rightful* (2), *riches* (1)
LEFT grammatical		RIGHT grammatical
and (39), *the* (10), *that* (14), *to* (13), *for* (10), *y* (9), *of* (13), *me* (9), *in* (6), *as* (8), *by* (9), *with* (5), *hit* (5), *his* (5), *you* (3), *him* (1), *a* (4), *all* (6), *not* (4), *or* (2)	*reason*	*and* (38), *the* (24), *that* (17), *to* (12), *for* (14), *y* (15), *of* (6), *me* (6), *in* (7), *as* (5), *by* (3), *with* (7), *hit* (5), *his* (5), *you* (6), *him* (8), *a* (5), *all* (3), *not* (4), *have* (8), *or* (1)

As observable above in Table 6.12, *reason* is shown to collocate with a selection of lexical and grammatical collocates. The lexical words provide a semantic representation of the thematic constructs linked to *reason* within the poem (*right/rightful, truth, conscience*). *Reason* also demonstrates a recurring textual proximity to lexical collocates containing the same word-initial phonemic sound /ɹ/ (*right, rich, read, ride, rule, rightful, riches*); this pattern extends to terms not shown in Table 6.12, which only collocate with the keyword in one or two instances: *rebuke, record, rather, reave, rehearsed, renkes, recorded, refusede, rape, rood, roamed, rode, rest, reuerensede, renoye, repentance,* and *reward.* When linked to *reason*'s demonstrated co-occurrence with coordinating grammatical conjunctions (*and, or*) it produces the lexico-grammatical construction of metrical alliteration, a linguistic construct characteristic not only of *PPCorp* but also of the alliterative revival texts in *ARCorp*.

One of the most notable patterns in the collocates for *reason*, aside from its personification-based fictional world functionality evidenced in patterns such as *reason said* and *reason rode*, is the highly technical nature of many of its lexical collocates. Out of the 15 lexical collocates identified in Table 6.12, 2 of them constitute other legal KWs generated for *PP* (*truth, conscience*) and an additional 5, while not key to *PP*, are identified by Alford (1988a) as lexis possessing technical legal significance within a 14th century context (*king, sire, right, rule, pity*). In addition to the combined technicality of *reason* and its various collocates in isolation, their textual use in *PP* in binomial constructs increases the technicality by forming phrasal constructs that themselves have legal significance beyond its separate parts. KW. For example, *right* collocates with *reason* a 25 times out of the 109 total occurrences of the KW in *PPCorp* (see Figure 6.13 below). Of these 25 occurrences, 6 are comprised of the binomial *right and reason* (15, 25, 38, 52-4). This binomial exists as a multi-word linguistic unit which, according to Alford (1988a: 137), can be traced back to a tradition of use within equitable law with a meaning associated with justice or equality in legal practice.

Figure 6.13: Concordance of the 109 total occurrences of *reason* in the annotated *PPCorp*, as generated in *WordSmith*.

N	Concordance
1	heles Captive other reneyed as hit seems As ac **reason** shall rekene with him and rebuke him at the la
2	Mede and men of ?our craft much truth lets Ac **reason** shall reckon with ?ow if y reign any while And
3	coest had such imperfect people reprieved all **reason** And halt hem vnstedefast for hem lakketh case
4	e hit made With guile you hem get a-gaynes all **reason** For in my palace paradise in person of an adde
5	stice Be ransomed for his repentance there all **reason** him condemn The There that party pursues the a
6	ere and of me y may the bet hem claim Al-thouh **reason** redordede and right of myself That if they ate
7	is The king to counsel tho took Conscience and **reason** And modiliche upon mede many times looked And
8	men that with the world deal That rich ben and **reason** knows rule well ?our souls Beth not unkind y c
9	he ploughman became so priueyliche he went And **reason** ran after and right with him ?ede Save Conscie
10	hen Conscience on his capel comes to pryke And **reason** with him ry?t rounynge together Which a master
11	e king to Conscience then Rape the to ride and **reason** that you search Command him that he come my co
12	r loves thy lore and leaves as you techest And **reason** aresonnede me and said reach the never why y s
13	s klikened to lewetee that last shall ever And **reason** to rich godlde to right and to truth The third
14	he field full of folk from end until other And **reason** y-reuestede right as a pope And Conscience his
15	d and by this light said y haled hit right and **reason** of my reave to take Al that my Auditour or els
16	r god gave salomon grace upon earth Riches and **reason** the while he right lived And as son as god say
17	y why ?e worden to me thus was for y aresonede **Reason** ?e certainly said he that is truth and shop hi
18	ated in my heart Repent the said repentance as **reason** the taught And shrive the sha?prpeliche and sh
19	and ignorant Ermytes For y made of tho men as **reason** me taught For as y cam by Conscience with reas
20	deliciis comeditis in tor euo & cetera Then as **reason** quickly anon right after That Conscience comma
21	atly Then any pore patient and that prove y be **reason** Hit are but few folk of this rich that nor fal
22	o wey where both mosten rest And rikene before **reason** aresonable acounte what one hath what a-noþer
23	y by Christ said Conscience congeie me are But **reason** read me þertyl rather would y day And y comman

ft the baldore to be to beat my servants Butte reason have pity on him he shall rest in my stokkes A
t A robs me of his mastery For by right and by reason þ the renkes that ben here Body and soul beth
pore and in rich And the rich is reverenced by reason of his riches There the pore is put behind and
each one Kings and knights should care hit by reason Rydon and rap a-doun in tides about And take t
so doth not other beasts They rule h.em al by reason Ac renkes full few And therefore merueileth me
And on hey heaven is priueoste and next him by reason And for the fairest fruit before him as of ear
hat the fend taught and his flesh will A?eynes reason y rihtwysnesse record hit with truth That her
ho-so takes a?eyn truth or transverses a?eyns reason Truth shall do him law and no life else Shall
clergy ca can & conceived more then reason For reason would have rehearsed the right as clergy said
hit were said the king well all about For-thy reason readily you shalt not ride hence But be my chi
g met Corteyslyche the king then cam and greet reason And between himself and his son set tho sire r
god he was pot out of bliss Right so happened Reason by the for thy rude speech And for you would w
lords allows him little or appeals ere to his reason A tempers the tongue to truth that no treasure
y shall say as y say sleeping as hit were How reason quickly al the tide to heaven He bad wastrels
ai To rule all tides therewith in right and in reason lo here the letter said he latin and hebrew Th
at you community defend And rewle thy realm in reason right well and in truth That you have al your
orabitur opus mersenarii & cetera And there is reason as a reave rewardynge truth That both the lord
rcy For unkindness is the contrary of all kind reason For there nor is pity nor sorrow nor non so mu
tera Hit were a?eyns kind said he and all kind reason That any creature should ability al except chr
And al the wit of the world was in three kings Reason and rihtfulnesse and pity they offrede Where-f
of You know well said Conscience and you kunne reason That knight king conqueror may be o person To
ved Pride or presumption of thy perfect living Reason refusede the and would not rest with the Nor c
d hit seems to all wits Right so said that man reason hit shows that he that knows clergy ability so
er speech Rehearse there anon right that might reason stop Ac al rightful recorded that reason truth
o as truth teaches Then is racio my right name reason an english And when y feel that folk tells my
en ?eueth before the doing And that is neither reason nor right nminime in no realm law That any man
e a saint and such a thef together Hit were no reason nor right to reward both y-lyche And right as
cristendom to renoye Rihtfolliche to renoye no reason hit would For may no cherl charter make nor hi
e help are needy pore And that is no right nor reason for rather me should Help him that hath not th
otherwise changed For-thy rebuke me right not reason y you pray For in my Conscience y know what Ch
nd even in the midst He may resceyue right not reason hit shows For the fyngeres that fold should an

his make . and thereof me wondrede For out of reason they ride and rechelesliche token on As in sec
With the castle of chest and chattering out of reason There counts count of courtesy and consents to
teach the to praise truth and take counsel of reason y have a knight Conscience cam late fro be-?en
follow hem and fast ride after To take reed of reason that record should Before the king an Conscien
asting days before none fed me with ale Out of reason among wicked people here rybaudrye to here Her
as himself Of lore and of letter of law and of reason so with that clergy can and counsel of scriptu
rs There the pore dare plead and prove by pure reason To have allouaunce of his lord by pure law he
e anger And be impatient in here penances pure reason knows That they have cause to oppose by kind o
g as y live for his luther works Some men read reason tho to have pity upon that shrew And for to co
ones And how that mede toke and mercy refusede Reason y say truthfully sew all beasts In eating and
em not . neither rich nor pore Then y resonede reason and right until him y said y have wonder in my
ed forth and with reason we met we reuerensede reason and roamed forth softly And meet witha master
dark to ben robbed Then he that is right rich reason bears witness Pauper ego ludo dum tu dives med
to rule the people Ac there the rote is roton reason know the truth Shall never flour nor fruit gro
thus even By him that rauhte up the rode said reason to the king But you rule thus all tides sorrow
ivy letters Nor no writ send but y assent said reason And y dare leg my life that love will lene the
all my Courts be a kings Justice y assent said reason by so yourself here Audiatis alteram partem am
n receiver Mede they besought Read me not said reason no pity to have Until lords and ladies louen a
e took his leave y shall array me to ride said reason rest the while And kalde Catoun his knave Cort
es as y see other Y say it not for myself said reason and it so were That y were king with crown cro
or priest other penitent for his sins And said reason by the rode y shall no pity have Whiles mede h
to duyren Then have you lands to live by said reason or lineage rich That find the thy food for an
my time to perfect shall turn Y read the said reason tho rape the to begin The l..if that is lovabl
gs treasure Conscience know well And also said reason y read ?ow rich And comuners accord in all kin
r fait lui suffrable a bien venir For-thy said Reason y read you rule thy tongue ever And are you la
caused me and my moed chaungede was that y say reason sewen all beasts save man and mankind many tim
e that he will not Reuen vs of our right sethe reason hem dampnede And affirm we have ben sesed seve
n And between himself and his son set tho sire reason And speak wise words along while together And
my womb one And also more-ouer me thinks sire reason Me should constrain no Clerc to no knaves work
Conscience that Courtesy serve Ride forth sire reason and rech not of here tales For there is wrath
and pore him loved And sayden we say well sire reason be thy words That meekness worth master over m

science his cross bearer before the king stand Reason reverently to-fore al the tide Preached and pr
ue sunt dei deo or ?e don ill For ri?tfulliche reason should rule ?ow all And kind wit be wardeyn ?o
id And kind and Conscience courteously thanked Reason for his rightful speech rich and pore him love
asses or sit and write Read and resceyuen that reason ought to spend Ac since bondsmen barnes have b
nscience know this for kind wit me taught That reason shall reign and tides govern And right as agag
might reason stop Ac al rightful recorded that reason truth said And kind and Conscience courteously
 dismiss all his officers And receive tho that reason loved and right with that y waked Passus quint
 our lords love other our lady gods moder Then reason rood forth Ac took reward of no man Butte deed
known that clergy ca can & conceived more then reason For reason would have rehearsed the right as c
ay room All these route of rat catcher to this reason they assentide Ac tho the belle was y-broughte
k not on him until him lust to rise For though reason rebuke him then cares heneuer Nor of clergy no
sts save man and mankind many times me thought Reason releuede hem not . neither rich nor pore Then
che by mastery But.e through right and through reason ransom here my leg Non veni soluere legem set
tt one me apposede Romynge in remembrance thus reason me aratede Can you serve he said or sing in a
 have power to pynschen hem then pot forth thy reason For y bykenne the christ said she and his clea
sterre Right so teaches letter ignorant men to reason And as a blind mana single minimman in battle
 law Ac Conscience knew hem well and carped to reason Here comes said Conscience that Courtesy serve
in faith tho said Conscience And rood forth to reason and rouned in his ere And said him as the king
hem called App tag clear? That a king cam with reason ykyuered under ensense The second king sees tr
y after And beden me rise up and rome and with reason should y dine And y aroos and roamed forth and
ason me taught For as y cam by Conscience with reason y met In an hot harvest when y had my hele And
d y dine And y aroos and roamed forth and with reason we met we reuerensede reason and roamed forth
cursed christian cattle and riches worschipede Reason and rihtfoldoem tho religious dampnede right s
ster was maed sit first As for the most worthy Reason stoed and styhlede as for styward of hall Pati

To this end, the functionality of *reason* in *PP* within a thematic signalling capacity is fundamentally linked to collocationally-conveyed levels of technicality. To further examine this trend, the following section introduces a linguistic framework for approaching technicality and applies it to a) the legal KWs generated for *PPCorp* before b) exploring the significance of *reason* and its collocates.

6.4.2 Technicality and legal language

As a highly specialised professional variety of language use, legal language benefits from analysis conducted based on parameters of technicality. Linguistic examinations of legal language are typically based on taxonomies distinguishing between levels of technical meaning like the one provided by Marin and Rea (2014) and Marin (2014). Building on a system originally created by Chung & Nation (2003: 105-6; 2004), this framework offers a non-temporally specific taxonomy for the classification of legal vocabulary using both qualitative (semantic) and quantitative (frequency-based) criteria. Given that present investigation does not involve the comparison of a legal corpus to a general reference corpus, any frequency-based parameters embedded in this framework are treated as guiding factors rather than statistical requirements, as set out below in (a-d):

(a) Words denoting a legal concept which are frequently used both in the general and specialised fields, not changing their meaning in the legal context.

(b) Words often employed both in the general and specialised fields which change their meaning in the legal context, sharing some semantic features with their original meaning.

(c) Words occurring more frequently in the specialised field than in the general one. These words change their meaning in the specialised context, their

new meaning being quite distant or completely unrelated to their general one.

(d) Highly technical words which appear almost exclusively in the legal field. If these words are employed in a general context, they also convey a legal concept in that field (Marin 2014: 57-8).

This listing provided by Marin (2014) maps legal language onto a cline of technicality from general to more specific. The first group (a) represents the least technical legal terminology, featuring words whose meanings remain the same despite the context of their occurrence and in doing so most clearly demonstrate the origins of legal vocabulary as stemming from common language use (Hall 1949: 37). These terms stand in constrast to the final category (d), which contains technical legal terminology possessing highly specific law-oriented meanings, or legal jargon. While words in both the first and final categories are characterised by their pragmatic consistency on opposite ends of the technical spectrum, lexical items in the latter category tend to be limited to use within legal contexts.

Of greatest interest to the current investigation are the middle two categories in Marin's (2014) framework, (b) and (c), encapsulating terms falling under the penumbra of sub-technical language (Cowan 1974, Flowerdew 2001). Unlike the field-specific technical terminology constituting jargon, sub-technical broadly refers to words 'in general usage, but which have a special meaning within the technical area' (Flowerdew 2001: 78, citing Inman 1978). The legal keyword *caes* (646), for example, represents a prototypical term in this category; it generally refers to a matter or affair, particularly concerning identified individuals or circumstances, yet also possesses technical meanings in a legal context as a contested question brought before a court of law, or a side of one party in a trial. It furthermore retains technical meanings outside of the law as an instance of a disease or the condition of a specific patient within a

medicinal context (*MED*). My idea of this sub-technical language combines (b) and (c) of Marin's (2014) taxonomy to gain a collective sense of this sub-technical terminology in the poem inclusively, irrespective of frequency measures. This decision closely aligns with the approach taken by Marin and Rea (2014: 66) in broadly defining sub-technical legal language as 'terms which have a different meaning when in contact with [a] legal context'.

Sub-technicality is a significant concept to the current investigation because, within the remit of the key legal language being examined in *PP*, terms that can be classified as sub-techical are most likely to be of stylistic potential. The multi-functional nature of these terms as being technical or non-technical indicates a high likelihood of polysemy within this category, which is confirmed in the classification of the poem's legal KWs according to Marin Perez's (2014) taxonomy, as set out below in Table 6.14. This categorisation organises the poem's legal keywords into groups of general, sub-technical, and technical terminology; general legal terms correspond with (a), with sub-technical legal words correlating with the modified sense of (b) and (c) and technical legal terms with (d) in the taxonomy put forth by Marin (2014).

Table 6.14: Classification of the *PPCorp* legal keywords by technicality

General legal terms		Sub-technical legal terms		Technical legal terms
law	*justices*	*conscience*	*rect*	*dei*
clerks	*bondsmen*	*mede*	*maken*	*ad*
writ	*sysores*	*ignorant*	*leely*	*symonye (simonye)*
leave	*indulgences*	*community*	*falseness (false)*	*nolite*
privy	*mayres*	*courtesy*	*justice*	*si*
guilt	*for*	*truthfully*	*shows*	*restitucioun*
movable	*reume*	*(sothfaste,*	*rightfully*	*presents*
possessions	*bulls*	*soeth, treuth)*	*issue*	*peccatum*
help	*me*	*reason*	*iouste*	*legem*
syuyle	*notaries*	*witnesses*	*conseyl*	*pena*
(syuile)	*burg*	*(witness,*	*(conseyle)*	*arrears*
god	*seal*	*wittiness)*	*release*	*nomine*
perpetual	*alawe*	*mercy*	*claims*	*culpa*
thieves	*commissary*	*saved*	*(claymen)*	*vsure*
reave	*bastards*	*catel*	*lordship*	*manibus*
curatours	*pauper*	*pardon*	*pays (payede)*	*iudicij*
(curator)	*fellowship*	*pees*	*sekte*	*debes*
faith	*faytour*	*loyal*	*allowed*	*reddet*
akyng	*y-maried*	*penances*	*borweth*	*quasi*
(kings)	*rihtwysnesse*	*grace*	*apose*	*non*
tithe	*raueschede*	*cam (comes)*	*caes*	
contreuede		*learned*	*euenesse*	
parsons		*amend*	*(forgiveness)*	
		lords (lord,	*places*	
		lordeyne)	*suffer (soffrede)*	
			doctor	
			knows	

6.4.2.1 Technical KWs

As shown above in Table 6.14, the technical legal terms in *PPCorp* constitute the smallest of the functional categories. The literary nature of the text makes this lack of highly technical legal terminology unsurprising; however, the dominance of Latin terminology in this category reflects the distinctiveness of this technical language within the texture of the poem (Alford 1977). The few English words in this category collectively refer to legal states of fiscal responsibility, with *arrears* (567) alluding to a

standing debt, *vsure* (651) to a standing loan made, and *restitucioun* (435) to the legal requirement to return unlawfully acquired money or property (1988a: 135). While possible to attribute the comparative absence of English in this category to a relatively small lexicon of highly technical English legal vocabulary in the late 14th century, the results provided above also enables us to observe that Langland chose to more frequently address legal topics in the poem using less technical terminology, such as those seen in the general and sub-technical keyword categories.

The overwhelming dominance of Latin in the poem's technical keywords invites further consideration of their textual role. While words including *legem* 'law' and *culpa* 'fault/responsibility' illustrate fundamental legal concepts on their own, the majority of terms are drawn from everyday language and appear embedded within recurring Latin expressions. The keywords *dei* 'day' (61) and *ad* 'to' (142), possessing fairly strong keyness values at 106.76 and 50.77 respectively, both assist in forming legal phrases such as *ad pristinum statum ire* 'to return to the first state' and *in dei nomine amen* 'in the name of God, amen' (Alford 1988a: 2; 72). *Non* (756), while just barely making the list with a keyness value of 15.32, operates in a similar fashion yet demonstrates a greater range of functionality than either *dei* or *ad* combined. It appears on its own in phrases such as *clemenica non constringit* 'mercy is not restricted', *ignorantia non excusat* 'ignorance (of law) does not excuse', and *non morabit[ur] opus mersanarii* (a maxim calling for prompt payment for work performed); the word additionally appears in phrases containing other technical keywords, such as *necessitas non habet* ***legem*** 'need has no law', and an abbreviated version of ***pena*** *pecuniaria non sufficit pro spiritualibus delictis* 'pecuniary penance does not suffice for spiritual faults'. These expressions have been identified as set

phrases with legal origins, either as maxims of canon[63] or natural[64] law, or derived from their formulaic use within earlier legal documents.[65] The use of *ad, non,* and *dei* within their corresponding legal phrases only occurs once for each keyword, making their legal functions fairly minimal in their overall textual impact as phrasal hapex legomena.

These technical keywords operate as a collective unit within the text. Very few possess particularly high keyness values, with the majority appearing towards the lower end of the keyword list (*peccatum, legem, nomine, culpa, manibus, debes, reddet, quasi*). Additionally, with the exception of grammatical words such as *non*, few of these technical keywords have very large raw frequencies. However, as a group these Latin terms collectively establish legalistic overtones throughout the text through their presence in these identifiable legal phrases, operating through embedded law-oriented associations derived from their original uses in legal Latin. While not limited to legal (or even technical) uses in Latin, the appearance of these terms interspersed within a piece of medieval English discourse generates this connotative legal specificity in their usage.

Like the majority of Latin references in the text, the phrases containing these technical keywords almost exclusively appear within long monologues within the narration or a fictional speech act, therein serving to add the weight of legal emphasis and power to an individual's argumentation. This is most well-recognised in the trial of Mede (Passus 2-4), when Simony first recites the standard opening lines of a charter of conveyance to lend legal credence to an otherwise ethically ambiguous engagement

[63] See Alford (1988a:2, 71).
[64] Ibid. p. 102.
[65] Ibid. p. 72 (wills), 113 (written law).

(vi), then in Conscience's technique of using legal maxims to strengthen his arguments in his debate with Mede (vii).

(i) ***Sciant presentes & futuri*** & cetera
Wyten and witnessen þat wonyen on erthe
That mede is maried more for here richesse
Then for holynesse oþer hendenesse oþer for hey kynde.

Be it known to all present and to come*, etc.*
Be it known and witnessed by all who live on earth
That Mede is married more for her riches
Then for holiness or nobility or for high birth. (2: 78-86)

(ii) y halde hym ouer hardy or elles nat trewe
That ***pre manibus*** is paied or his pay asketh…
And Gylours gyuen byfore / and goode men at þe ende
When þe dede is y-do and þe day endit
And þat is no mede but amercedeamanere dewe dette
And but hit prestly be y-payed þe payere is to blame
As by the book þat byt no-body with-holde
The huyre of his hewe ouer eue til amorwe
Non morabitur opus mersenarii & cetera.

I hold him overconfident or else not true
That is paid ***before hand*** *or asks for his pay [in advance]…*
And cheaters pay before and good men at the end
When the deed is done and the day ended
And that is not mede but merit and due reward
And but if not paid promptly the payer is to blame
As by the book that bid no one withhold
The wages of his workers overnight until the morning
The wages of him that has been hired shall not abide with you until morning, *etc.* (3: 292-310)

What is interesting to observe about the use of these highly technical legal phrases is their coordination with items of vague language such as 'etcetera'. A type of grammatically conveyed extender in English (Overstreet 2015), the term typically appears as the closing element of an enumerative sequence (Cortes Rodriguez 2006) operating as an indefinite allusion to further unspoken information or examples related to already articulated concepts. In *PP*, '& cetera' is a particularly repetitive

textual feature used in conjunction with these legal Latin phrases, a correlation evident in (i) and (ii). In the case of (i), the use of '& cetera' in *sciant presentes & futuri & cetera* belongs to the phrase itself, functioning as a traditional component of the standard opening line of a deed or charter of conveyance (Alford 1988a: 140). It can also be used in the text to shorten a lengthier phrase through textual substitution of its final elements, as seen in *pena pecuniaria non sufficit [pro spiritualibus delictis]* (1988a: 113). Its abbreviating function here serves a practical purpose in enabling a degree of brevity for the poem's intra-linear metrical and alliterative constraints, but it is also indicative of a higher pragmatic function as well. The abbreviation of the legal phrase implies a degree of mutual understanding between interlocutors, indicating that the speaker has confidence that those listening to his argument will a) recognise the phrase, and b) understand its use and the surrounding legal significance attached to it. This function is also apparent in (ii), albeit for different reasons. As an aspect of canon law, *non morabitur opus mersenarii* is not traditionally linked to the use '& cetera', indicating that its presence here is not to finish the rest of the legal phrase, but rather to imply an unspecified list of statutes within canon law that support the opinion that workers should receive prompt payment of their wages (1988a: 104). The unspecified nature of this list can again be a stylistic choice by Langland not to burden his verse with lengthy recitations of canon law, but could also indicate the degree of contextual knowledge shared between discourse participants, wherein continuing with such a list would be unnecessary based on the audience's existing knowledge of the unmentioned legal subject matter.

6.4.2.2 General KWs with legal applications

The keywords appearing within the general linguistic category (the 'general legal terms' category in Table 6.14) represent legal terminology in the poem whose meanings remain constant both in legal and everyday contexts. Prototypical keywords in this group include common concepts such as the *law* (40) and ownership through *possessions* (198), and tend to occur more frequently in the text than the majority of words appearing in the technical legal category. The largest group of terms is comprised of descriptors, which characterise individuals based on profession (*clerks, notaries*), criminal activity (*thieves*), social position or status (*bondsmen, lords, kings*), marital status (*y-maried*), truthfulness (*faytour*), and legitimacy (*bastards*). The legal meanings of these descriptors are found embedded in their general definitions and are frequently used in general contexts, subsequently carrying these legal associations into every use of the term with each occurrence. This implicit legal association is also present in terms such as *guilt* (195) and *symonye* (145, 584), which possess very strong legal connotations within both legal and everyday communicative contexts.

In addition to embedded legal associations, a number of these general KWs convey legal meaning only when linked to legal subject matter. For instance, Figure 6.15 below shows how the word *perpetual* (241) broadly refers to an unlimited span of time in describing the continuance of peace in three of nine occurrences in the text (ix, xiii, xiv), demonstrating its more law-oriented functionality only through references to legal sentences without set limits (Alford 1988a: 114). Two of these legal uses allude to the continuous pain suffered as punishment for alleged wrongdoing (xi, xiii), with another applying this timelessness to pardons given to reformed individuals (xv). These legal functions additionally adopt more stylised approaches to pardon and punishment

through characterisation via emotion (*perpetuel blisse*) (viii), experiential states (*pain*) (xi), as well as reference to theological constructs (*perpetuel helle*) (x).

Figure 6.15: Concordance of the nine occurrences of *perpetual* in the annotated *PPCorp*

(viii)	pain passed forth patiently to	**perpetual**	bliss Then marie magdalene wh
(ix)	here should be plente and pees	**perpetual**	ever Wedur-wyse shipmen now a
(x)	t Purgatory for thy pay Ore Or	**perpetual**	hell For shall no pardon pray
(xi)	it with truth That her pain is	**perpetual**	no preyer may hem help For-th
(xii)	sorrow and lucifers fythele To	**perpetual**	payne or purgatory awikke For
(xiii)	tients vinciunt vincunt Before	**perpetual**	pees y shall prove that y sai
(xiv)	of heaven descended To make a	**perpetual**	pees between the prince of he
(xv)	r lord truth hem grants pardon	**perpetual**	right as piers the ploughman
(xvi)	ven Three persons indepartable	**perpetual**	were ever Of o will of o wit

Additional legal keywords in this category also include commonly used words such as *help* (213)*, god* (233)*,* and *me* (551), which belong in formulaic English legal phrases such as the traditional witness' oath *so me god help* (1988a: 146).

Rather than adding overt legal overtones and authority to the poem as done with the technical keywords, these general legal terms tend to reveal broader socio-cultural conditions and concerns featured within late medieval English law. The nine repeated collocations of *possessions* to *movable*, which also appears as a general legal keyword, highlights the medieval necessity to legally acknowledge instances where ownership extends beyond grounded property (i.e., lands and buildings) to items movable by a person (monetary wealth) or of their own accord (livestock) as *mebles possessions*. The textual use of *law* itself in *PPCorp* reflects this historical contextualisation in the different types of law present in late medieval England through reference to *commune law*, *Westminster law*, and *God's law,* the latter also more creatively described as *his law* and the *law of Holy Church*. The additional use of *old* and *new law* to refer to rules governing the ecclesiastical courts also reveals the broader ideological changes taking place that saw the law of old Christian narratives

being replaced by new scriptural texts and beliefs. The presence of these legal phrases provides not only an interesting perspective on the perceived amount of overlap between the various types of medieval law in the governance of late 14th/early 15th century daily life, but ongoing changes within these overlapping legal domains as well.

While this overview of the technical (6.4.2.1) and general legal keywords focuses on prototypical words, these two groups also feature somewhat less paradigmatic examples, particularly in the general language category. The generated KW *for* (514) reflects this categorical difficulty in overlapping between the general sub technical language groups by virtue of representing three separate legal terms, each linked to a different legal concept. Out of its 1,168 occurrences in *PPCorp*, *for* functions as part of the compound phrase 'for ever' 14 times; like *perpetual,* some of these uses relate to punishment, but additionally extend to the possession of landed property. The keyword also contains five instances of the legal concept *forgiveness* created by merging the separately generated keywords *for* and *euenesse* (343) to refer to the official cancellation of an existing debt. Use of the term in the poem similarly includes three occurrences of *forgive* as an action of pardon for a crime or personal offense (Alford 1988a: 63). In such instances, where classification is divided between different functions within a keyword entry, the keyword has been placed in the category most frequently utilised in the text as determined by its use in at least half of its concordance lines. This approach accounts for why the keyword *for* appears in Table 6.14 as a general legal term, while *forgiveness* and *euenesse* have been placed in the sub-technical KW category.

6.4.2.3 Sub technical language

As demonstrated in Table 6.14, the majority of legal keywords present in *Piers Plowman* appear as sub-technical language. These keywords represent terms observed to be closely related to their general meanings in everyday communication, but demonstrate differences in pragmatic scope and additional semantic nuance when utilised for legal purposes. The keyword *catel* exemplifies this separation between general and sub-technical legal terminology, generally referring to livestock but in its legal sense broadening to encompass all types of movable goods or possessions as protected by the law (Alford 1988a: 23). This semantic distinction also extends to a number of verbs present in the sub-technical group (*cam, amend, saved*). While addressing the actions of arrival, rectification, and protection reflected in common usage (*MED*), these terms also possess narrower meanings within legal contexts; *comen* specifically refers to an individual's appearance in court (1988a: 30-1), while *amenden* and *saven* both allude to redemption of wrongdoing or imprisonment through legally-mandated financial restitution (1988a: 5; 139).

These sub-technical keywords address several closely related themes within the poem. A number of terms involve the act of reparation, be it financial (*mede, saved, amend, forgiveness*) or confessional (*knows, forgive*). This thematic focus on making amends – and the benefits of doing so – constitutes one of the primary focuses of the text (Simpson 1990), and is intrinsically tied to those keywords relating to wealth distribution (*makes, pays, tithe*) as an aspect of this reparation process. The legal aspect of making amends as an obligation (Galloway 2011: 372-3) also links other sub-technical KWs relating to the courts through proceedings (*comes, claymen, apose*) as well as the persons traditionally involved in a reparation case (*sekte, witnesses,*

conseyl). Rather than approaching the concept through the legal lens of reparation, a number of literary scholars (White 1988, Pearsall 2009) refer to the 'salvation' of amends, and in doing so divert discussion away from the legal and towards more theological subject matter, enabling parallel discussion regarding atonement as both a spiritual and legal process.

Categorising legal terminology based on technical parameters thus provides a method of studying the range of functions present in sub technical KWs such as *reason* as both legal and non-legal and, when distinct from one another, offers insight into their legal versus poetic functions within texts like *PP*. Ascertaining these functions is based on the amounts of technicality conveyed in the collocational patterns surrounding the use of the word in the poem, which builds on the technicality embedded in the collocates themselves initially observed in 6.4.2. For example, the high amounts of collocation of *reason* with grammatical coordinators such as *and* and *or* signposted in the L1 and R1 positions suggests that the KW recurringly appears within binomial constructs in the poem. This is confirmed in the KW concordances listed in Figure 6.13 in the appearance of phrases such as *right and reason* (15, 25, 38, 52-4), *conscience and reason* (7), and *riches and reason* (16).

From amongst these binomials, *right and reason* stands out as the one containing the most technical legal meaning. The construct exists as a multi-word linguistic unit which, as Alford observes, can be traced back to a tradition of use within equity law with a meaning associated with justice or equality in legal practice (1988a: 137). An example of the poem's use of this legal phrase occurs towards the end of Passus 21, where an unnamed lord incorporates the phrase into his oral argumentation concerning his right to collect rents from his tenants:

(lxiii) Thenne lowh ther a lord and bi this lihte saide
y halde hit **riht and resoun** of my reue to take
Al þat myn Auditour or elles my styward
Conseileth me bi here a-counte and my clerkes writyng
With spiritus intellectus they cote þe reues rolles
And with spiritus fortitudinis fecche hit wolle he ullhe. (21: 459-64)

Then a lord laughed there and said 'By these lights
*I hold it as **right and reason** of my reeve to take*
All that my auditor or else my steward
Counsels me by their accounts and my clerk's writings
With the spirit of intellect they took the reeve's books
And with the spirit of fortitude I'll fetch it whether he likes it [or not]'.
(21: 459-64)

The use of the binomial in this instance serves as one of several argumentative techniques employed by the character, who also incorporates Latin into his discourse as a method of exploiting the legal power of Latin in this pseudo-legal debate in which he finds himself a participant. While spoken by a fairly minor character, the inclusion of the phrase adds additional legal meaning to the situation through attempts to distance himself from legal guilt by assigning the burden of responsibility for accuracy in the accounts to his stewards, and also by representing his tenants' intentions as 'deceit and self-seeking masquerading as prudence' (Pearsall 1994: 360).

The phrase *right and reason* belongs to a larger group of binomials evidenced within medieval legal discourse. Operating as phrasal elements or individual words, binomials represent a defining feature of legal language in both spoken and written forms (Mooney 2014: 24). Phrases such as *in whole and in part* and *the freehold conveyed or long lease granted* are utilised in written legal language as structures which embrace the formality and ritual prioritised in the production and repeated use of legal formulae, while *to have and to hold* channels the more performative aspects of spoken legal traditions (see Tiersma 1999). Single-word binomials such as *property and chattel* are particularly frequent in early English legal discourse, and comprise the

majority of binomials found in medieval texts (Kopaczyk 2013). These three-word binomials often contain lexemes possessing different etymological origins (from a Middle English perspective, generally an English/French combination). Crystal (2003) and Schane (2006), among others, attribute this to a translation-based function which highlights broader concerns in using the right legal term for a given situation, an issue resolved by simply utilising both. In other words, making use of separate terms possessing different historical origins to describe the same referent enables more comprehensive legal coverage than selecting one of the words, particularly if the two terms are not absolute synonyms (Mooney 2014: 25).

In addition to their use in legal discourse, alliterative binomials also feature prominently in early English literary traditions. Kopaczyk (2013: 67) observes how 'the stylistic value and repetitive structures was recognised and exploited' in medieval literature, its origins – like law – similarly drawn from its use within an earlier culture of oral transmission. Rather than being derived from concerns of legal coverage or referential specificity (Koskenniemi 1968: 78), these more literary uses of binomials tended to focus on their value as extended mnemonic tools, particularly in Old English texts (Danet & Bogoch 1992). The practical functionality provided by alliterating binomials also corresponds with recognition of their aesthetic appeal; the combination of alliteration with the binomial structure creates an overarching coordinated parallelism exhibited on phonetic, syntactic, and semantic levels simultaneously (Damova 2007: 34-8). The use of these alliterative binomials has been shown to produce 'a powerful welding effect' which increases the 'strength and appeal' of the topics being addressed (Malkiel 1959: 122). Further analysis of this associated rhetorical power orients the study of binomials around documentary intersections of the legal and creative, where binomials as a recurring linguistic pattern would be

expected to prominently feature.[66] Within late medieval literary discourse, texts spanning across both the alliterative revival and the more law-oriented medieval debate poetry would match this profile, creating a shortlist of poetic texts of interest including *Wynnere and Wastoure, Sir Gawain and the Green Knight,* and *Piers Plowman.*

Given this information, it is perhaps unsurprising to acknowledge the presence of multiple alliterating binomials within the legal KWs generated for *PPCorp*. In addition to *right and reason*, the poem contains further instances of binomial use linked to these key legal terms, as shown below in Table 6.16. It bears noting that this list does not comprehensively cover all binomials in the poem derived from legal formulae, only those containing the text's keywords.

Table 6.16: Binomials containing one or more legal KWs from *PPCorp* (KWs denoted in bold)

right and ***reason***	
lered *and* ***leweth***	'learned and ignorant'
king *and* ***counsel*** *his crown* ***commune***	'king and community'
logic and ***law***	
love and ***leve***	'love and leave'
license and ***leve***	
moebles *and* ***unmoebles***	'movable and unmovable'

Within these binomials, the semantic properties of their individual elements reveal the nature of the relationships underpinning their coordination. The phrase *logic and law* represents a form of complementation in its reference to subjects

[66] While not limited to focus on binomials, Steiner and Barrington (2002) and Steiner (2003) examine this demonstrated constructive and discoursal similarity between the development of ME literary and legal texts

traditionally studied together in law (Alford 1988a: 89). This complementation also occurs in *king and his crown*, which distinguishes between the king's personal and official identities. Interestingly, the other two binomials beginning with *king* feature an additional hierarchical structure not seen in the complementary constructions. This overarching semantic hierarchy highlights the underlying power dynamics between each element, with *king and counsel* coordinating the common courts and their leading official (the king), and *king and commune* linking an authority figure with his subjects (1988a: 78). By contrast, *love and law* reflects antonymy in the dichotomous opposition of two different types of legal settlement, either through amicable resolution outside the courtroom (*love*) or formal litigation (*law*) (1988a: 91). *Right and reason* and *license and leve* are illustrative of what Kopaczyk (2013: 188) refers to as 'binomials proper', which contain elements exhibiting near-synonymous semantic repetition. These are the binomials, perhaps more so than the other types, which tend to be comprised of elements possessing different etymological origins; the uses of *license* and *reason* within their respective binomials derive from nominal uses of each term borrowed into English from French and Anglo-French respectively, while both *leve* and *right* are shown to have Germanic origins (*OED*).

What makes the use of these binomials striking in *Piers Plowman* is how these recognisably legal phrases are creatively incorporated into the text. Mollin (2012: 101) observes that binomials are largely formulaic, proceeding in a specific ordering based on internal semantic relations (as discussed above) as well as issues raised in iconicity such as chronological order and perceptual markedness (citing Malkiel 1959: 154). This fixedness – and the recurring use of these unchanging phrases – makes binomials recognisable, and has remained the focus of modern corpus linguistic studies of the constructions in English (see Mollin 2014). However, in looking at the concordances for

these binomials in *PPCorp*, it appears that *PP* contains a large number of textual instances where these binomial structures have been subverted. The single textual occurrence of the straightforward legal binomial *right and reason* in Figure 6.13 (15) stands in opposition to at least eleven distinctive permutations of that binomial within the text. Two occurrences, by *right and by reason* (25) and *through right and through reason* (98), essentially maintain the basic structure of the original binomial with only the addition of extra prepositional elements; this is most likely derived from syntactic and metrical considerations, rather than indicating any stylistic significance.[67] Five other instances are shown to introduce negation into the binomial with the substitution of *nor* or *not* for *and*. Three preserve the same ordering of its core elements (52-4), while the other two change the ordering of the binomial to *reason nor right* (49, 50). Compared to earlier alterations, this rearrangement of the word order is more indicative of stylistic choice in the ordering of the line's alliterative elements. This new binomial patterning is also seen in the use of the related but ultimately new constructions *reason and rightfulness* (43), *a?eynes reason and rihtwysnesse* (30), and *reason and rihtfoldoem* (108).

These creative adaptations of a recognisable legal binomial makes these uses of particular interest from a stylistic perspective. However, given the present information, it remains unclear whether this stylised use of recognisably legal binomials is a feature unique to Langland's work or something more prevalent within late medieval literary texts. As small and relatively self-contained packages of linguistic information, these binomials embody free-standing linguistic units able to be transferred across different literary texts with relative ease but this may not necessarily be the case with the more stylised versions found in *PP*. This point is

[67] See also *ri?tfuliche reason* (98 in Figure 6.13).

further considered in 6.4.3, using *right and reason* as a case study for exploring the greater utilisation of stylised legal constructs within medieval poetic traditions.

6.4.3: PPCorp vs. MEPoetLitCorp

This section starts with a look at *reason* and its functions in the poetic reference corpus. These late medieval poetic uses of *reason* are subsequently represented in a semantic lexical constellation, as put forth by Marin & Rea (2014: 68) in their study of legal sub-technical terms. A lexical constellation provides visualisation of the interconnected – and often overlapping – meanings of *reason* in *MEPoetLitCorp* and enables commentary on the differences in the pragmatic use of *reason* between *PPCorp* and *MEPoetLitCorp*, particularly regarding its legal functions. This focus on its legal meanings extends to discussion of the legal binomial *right and reason,* as well as exploration of more creative adaptations of the binomial in either the rearrangement of its core elements, or fundamental changes to its underlying grammatical structures.

Out of the 28 texts in *MEPoetLitCorp, reason* occurs 295 times compared to the 109 instances in *PPCorp*. Normalising these raw values to frequency per 10,000 words provides a more equal basis of comparison by which to observe the relative use of this term within each corpus (McEnery & Hardie 2012: 49-50). This reveals that *reason* occurs 15.4 times every 10,000 words in *PPCorp* while only occurring 3.3 times in *MEPoetLitCorp*. These normalised frequencies mean that for each use of the word in the poetic reference corpus, the word is used an average of 4.6 times in *PPCorp*. The significance of this difference in use was already revealed in the presence of *reason* as one of the top lexical keywords in *PPCorp* (45), but offers little information regarding

its functionality across the two corpora, which are more visible through its concordance lines.

Similar to its use in *PPCorp*, *reason* in *MEPoetLitCorp* broadly refers to the concept (or use of) logic as a higher cognitive process. It is used in nominal form both in the general intellectual capacity for rational thought (lxiv), or more specifically as a statement or fact used to support an idea or an argument being made (lxv). This narrower sense also extends to grounds for justifying an action or decision *by reason* (lxvi). *Reason* also operates in some instances as a descriptor for sensibility-based reasonableness (lxvii), and also as a broad reference to a healthy mind untouched by mental disease or illness (lxviii).

(lxiv) He was with love unwise constreigned,
And she **with reason was restreigned**:
The lusts of his heart he suieth,
And she for dred shame eschuieth,
And as she should… (*CA*)

(lxv) She said, 'pees! now take keep every man,
And herkeneth which **a reason I shall bring** (*PF*)

(lxvi) And thus by report was her name y-shove
That, as they grow in age, wax her love;
And certain, **as by reason of her age**,
There might have been bitwix hem marriage,
But that her fathers would not hit not assent. (*LGW*)

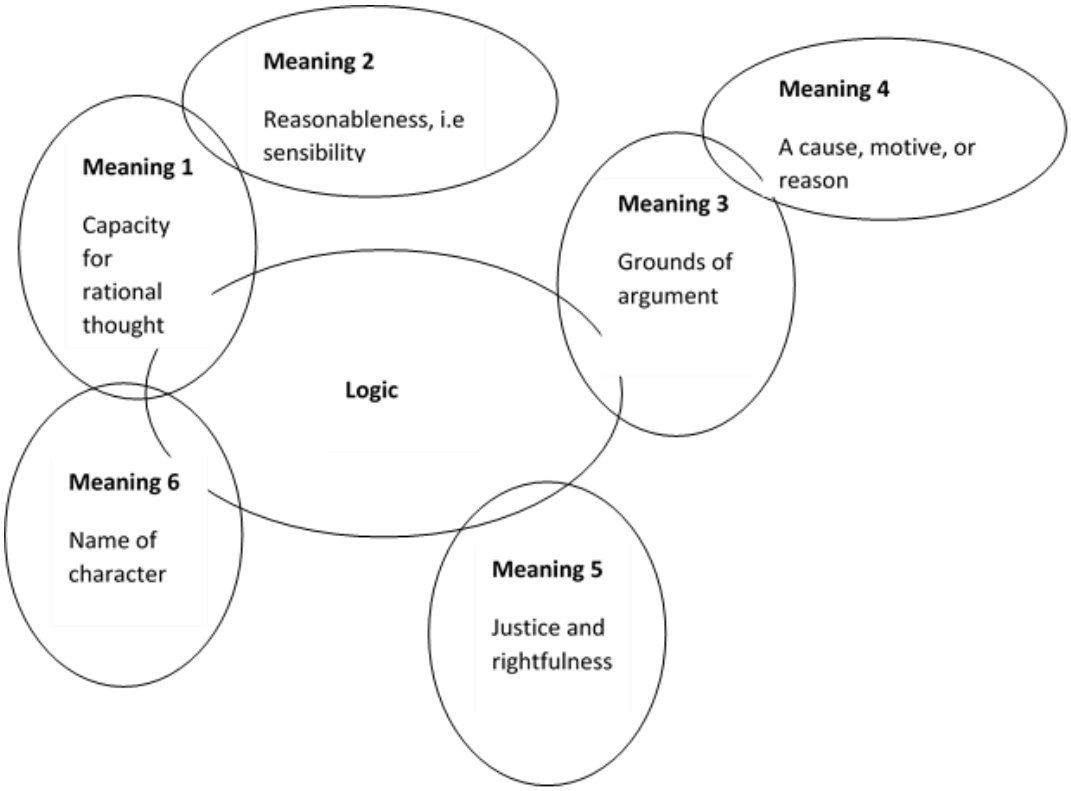
(lxvii) The truth is this, the cut fell to the knight,
Of which full blithe and glad was every wight;
And tell he most his tale, **as was reason**,… (*GP, CT*)

(lxviii) And she answered, 'sir, what eyleth you?
Have patience, **and reason in your mind**,… (*MerT*)

These uses of *reason* in *MEPoetLitCorp*, and their comparative frequency of use relative to one another, are visualised below in Figure 6.17. The centre (or nucleus) contains the core meaning of the word in a late medieval context of use, with related

meanings surrounding this core. Meanings containing a close relationship (a narrower example of a broader use) are represented by overlapping circles branching out from the centre of the diagram. Each meaning is numbered according to the frequency of its use within the 295 occurrences of *reason* in the reference corpus, based on study of its collocates and their concordances.

Figure 6.17: Semantic lexical constellation of *reason* in *MECPoetCorp*.



As shown in Figure 6.17, the majority of occurrences in the reference corpus allude to reason as a cognitive power signifying a person's ability to engage with the world in a logical manner. This is perhaps the most literal uses of *reason,* showing people as either possessing, demonstrating, or lacking reason entirely; common collocates signifying this use include *all reason*, *with/against reason*, and most coordinated nominal phrases ending with *and reason* to refer to reason as a general concept. Meaning 2 represents an extension of this intellectual ability, shifting focus

from the capability for reason itself to the degree of sensibility exhibited by the individual.

The next most frequent use of *reason* (Meaning 3) lies in expressing the grounds for a theory, opinion, or argumentative stance, with the majority of concordances occurring in the context of either legal, theological, or philosophical debates. This category functions to provide justification for the application of logic and reason for specific purposes, a function characterised by patterns such as *by reason* and its collocational association with terms such as *proof/prove* and *show(s).* This category also contains instances wherein *reason* appears as a predicate in phrases, resulting in the presentation of reason as a disembodied subject in passive constructions such as *reason is/was*. Such uses generally depict reason as a principle, one utilised in discussions addressing philosophical aspects of nature, or stances on topics of morality. Meaning 4 in Figure 6.17 operates as a narrower sub-category of Meaning 3, having a shared functionality oriented around logic-based justification; common patterns surrounding its use involve the use of possessive pronouns *my reason*, as well as definite and indefinite articles (*the/a reason for*).

Meanings 5 and 6 represent the least frequent examples of the use of *reason* exhibited in the reference corpus. The first category contains instances wherein *reason* appears in an evaluative sense assessing the rightness of a given situation. This focus on fairness orients this meaning of the term around judgement on the presence, degree, or absence of justice, and in doing so encompasses the majority of legal uses set out in 1A of Alford's (1988a: 134) definition of *reason.* Common collocates of the term in this particular sense include the majority of legal keywords identified in *PPCorp*, as well as any collocational patterns stemming from the association of these words (i.e, legal binomials). In contrast with the other categories, Meaning 6 is less a

change in meaning as it is pragmatic function. While taking on a new (and ultimately creative) dual metaphorical and referential functionality as the name of an allegorical character, the underlying meaning of *reason* remains unchanged. This constancy of meaning is reflected in the overlapping placement of Meaning 6 in relation to Meaning 1 and the core meaning of logic, demonstrating that these personified examples of *reason* borrow their fundamental meaning(s) primarily from these two categories, despite having different textual functions.

This semantic lexical constellation of *MECPoetCorp* thus illustrates the most prominent uses of *reason* observed in our representative corpus of late medieval literature. While not an authoritative (or even complete) mapping of the word in a 14th century context, the results here echo observations made in the literary criticism on the topic of reason, as discussed by Rudd (1994: 45):

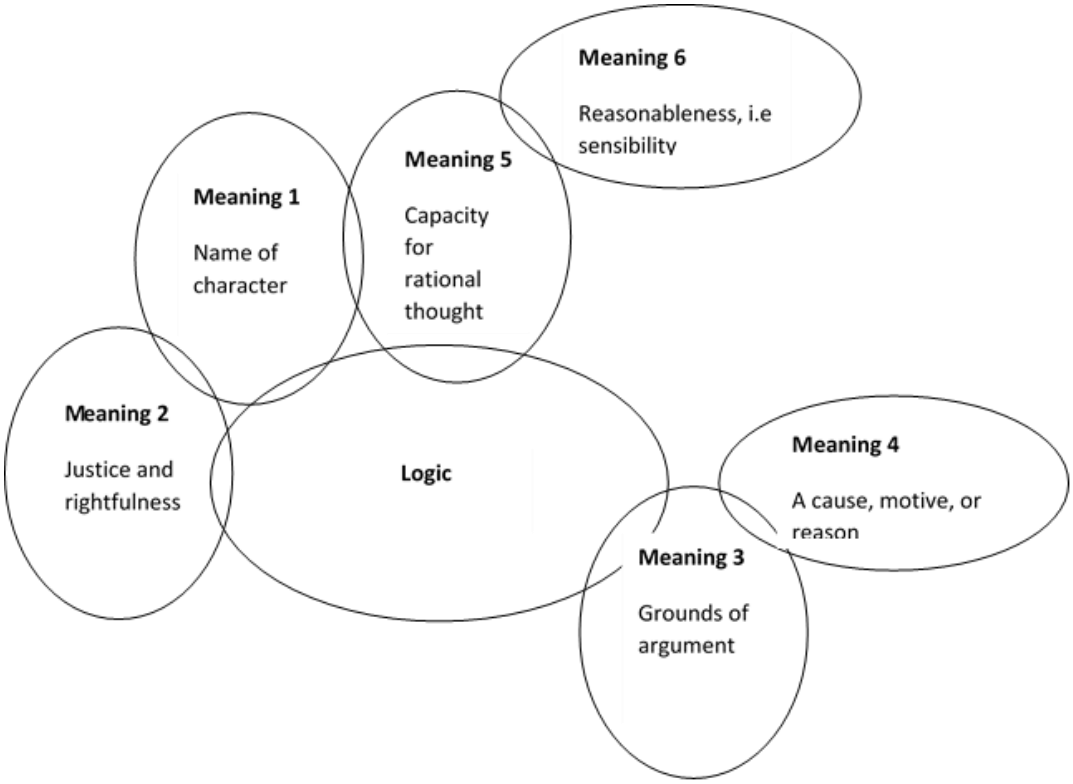
> Reason itself is a concept which attracted, and still attracts, much debate. For some it was the attribute which divided mankind from the animals and enabled humans to have some apprehension of the Divine; for others it did indeed assist in gaining such apprehension, but had to be left behind if the pinnacle of understanding was to be achieved; others again were more interested in the aspect of reason as the ordering and judgement faculty.

Earlier analyses of the use of *reason* in a late medieval context have traditionally focused on this idea of reason as a cognitive faculty, partially as a way of distinguishing its meaning from the use of *kind wit* as nature or a representation of natural reason.[68] This orientation gains context in light of the data drawn from our reference corpus, which shows Meaning 1 to be the most dominant meaning of *reason* across my corpus of late medieval texts. New discussions of *reason* as a concept have since extended to include acknowledgement of reason as a principle (White 1988: 7), particularly as one

[68] See Dunning (1937: 39-40), Baldwin (1981 :21-2).

responsible for the ordering of the universe (Rudd 1994: 44). This is largely due to work done by Alford (1988b) in highlighting the legal meanings of *reason*, in some instances overlapping with the argumentative grounds category (Meaning 3) to which this idea of reason as principle belongs. His work is largely based on the use of *reason* in *PP*, which shows a different semantic picture from the rest of the reference corpus, as seen below in Figure 6.18.

Figure 6.18: Semantic lexical constellation of *reason* in *PPCorp*.



As shown above, the frequencies ascribed to each meaning of *reason* in *PPCorp* are markedly different from those appearing in *MEPoetLitCorp*. Unlike the reference corpus, the majority of *PPCorp*'s uses of *reason* appear as personification used to represent the allegorical character Reason, as delineated by its presence as Meaning 1 in the new semantic constellation. Meaning 2 similarly changes between the corpora, becoming the legal use of *reason* as associated with justice and evaluative rightfulness,

which was previously represented by *reason*'s capacity to convey descriptive reasonableness in Figure 6.17. Both Meanings 1 and 2 in *PPCorp* undergo a radical change in comparison to their use in the reference corpus texts; both are represented as Meanings 6 and 5 (respectively) in *MEPoetLitCorp*, elevating these meanings of *reason* from the lowest-frequency categories in the reference corpus to the highest-frequency categories in *PPCorp*. Meanings 3 and 4 in *PPCorp* remain the same in their reference to the underlying causes, grounds, or reason(s) for either discussion or formal argumentation. Meanwhile, Meanings 5 and 6 as the intellectual capacity for reason and reasonableness appear as the least frequent uses in *PPCorp*, a demonstrable change from their very frequent uses within the poetic reference corpus.

One of the most notable differences between the two corpora is the number of times *reason* is used for allegorical purposes. It appears just under half of its total instances in *PPCorp* (51 out of 109) as definitive non-ambiguous personification, while only 34 of the 295 occurrences in *MEPoetLitCorp* possess the same textual function. While several texts containing allegorical figures exist across the reference sub-corpora – from *Parliament of Fowls* (*PF*), *Book of the Duchess (BD*), and *Romance of the Rose* (*RotR*) in *MVCorp* to *Pearl* (*Pe*) and *Wynnere and Wastoure* (*WW*) in *ARCorp* – only one of these texts contains a main character named Reason (*RotR*). Furthermore, the female character of Reason in *RotR* as an advisor of courtly love is not as central a character as the corresponding male representation of legal arbiter, moral advisor, and royal counsellor present in *PP*.

It is this distinction in the social roles played by these fictional characters which accounts for the constellation rearrangement in Figure 6.18. The role of advisor enacted by Reason in *RotR* is not advisory in a legal sense but rather in the capacity as

an independent arbiter, and in this sense operating as the embodiment of rationality personified. Because of this, the allegorical function of *reason* in the reference corpus is only linked to this capacity for reason represented in Meaning 1 (see Figure 6.17). By comparison, Reason in *PP* performs several roles simultaneously over the course of the text; his role as moral advisor through his debate with Will in Passus 11 and his sermon to the Field of Folk in Passus 5 is similar to that of Reason in *RotR* as the personification of rational thought. However, his role as counsellor to the king in Passus 4 provides additional meaning not seen in the reference corpus, and ascribes this allegorical function with additional legal meaning through his professional position as a royal advisor. From amongst the texts sampled for this study, this dual positioning of *reason* as a character is unique to *PP*, a feature whose presence is reflected in the overlapping positions of both the rational (Meaning 5) and legal meanings (Meaning 2) depicted in Figure 6.18.

By contrast, the justification-based uses of *reason* in the reference corpus are rather similar to those observed in *PPCorp*, albeit with some changes in scope. It appears as both Meanings 3 and 4 in both corpora (argumentative grounds and causes or motives, respectively), and are particularly observable in structures of oral debate, as seen in the humorous debate between two birds seeking a mate in *PF* (lxix). In these contexts, reason is utilised as the underlying groundwork framing the way each party sets forth its arguments; while closely mirroring legal procedures, neither the topic of discussion nor any of the individuals involved are associated with the legal profession. This framing of argumentation can also be seen in *House of Fame* (*HF*), as well as in the framing efforts set out in the prologue to *The Plowman's Tale* (*PT*).

(lxix) The tercelet said than in this manner:
Full hard were hit **to proof hit by reason**

> Who loves best this gentle bird here;
> For each one hath such replicacioun,
> That noon by skills may be brought a-doun;
> I can not seen that arguments avayle;
> Than seems hit there most be battle.' (*PF*)

The legal meaning of *reason* is also used similarly across both medieval poetic corpora. While evident that *reason* operates in a legal sense more frequently in *PPCorp* than the reference corpus, the collective majority of its uses in a legal capacity involves the invocation of the principle of justice and right order, particularly when characterising regional governance. The ability to rule *by reason and right law* or *rightful dome* occurs several times in *Mum and the Sothsegger* (*MS*), and in doing so follows the broader scope of the text in being a mediation on the merits of being silent or speaking out against contemporary institutions of state, including the legal system (Dean 2000); this standard of ruling in accordance to a sense of righteousness also features in discussions within *The Destruction of Troy* (*DT*). It also features in discussions of ecclesiastical topics, wherein one of the fundamental tenets of the faith establishes *the reason of right* in the Christian doctrine enabling the automatic acceptance of innocents – encompassing both young children and the mentally impaired – into heaven, as discussed in *Pe*.

This connection between reason and right parallels the continued observed use of these terms in binomial constructions within *MEPoetLitCorp*. Interestingly, there are no examples of the legal binomial *right and reason* as identified by Alford (1988a: 137). There are other binomials wherein *reason* functions as the final element, but these largely address non-legal topics which are localised to specific texts in the reference corpus. For example, the binomial *wit and reason* occurs seven times in the corpus, all of them in *Confessio Amantis* (*CA*). *Truth and reason* comes the closest to the legal

meaning of *right and reason*, but as an evaluative stance along the lines of veracity rather than fairness; it occurs twice in *MS,* and once in *CA.* Several other binomials following this structuring occur in the reference corpus, albeit only once each: *desire and reason* and *mind and reason* in *Troilus and Criseyde* (*TC*)*,* and *love and reason* again in *CA.*

Despite the lack of any examples of the legal binomial, the reference corpus does contain a range of binomials containing the reversed ordering *reason and right.* Figure 6.19 below provides a full listing of the concordances containing this binomial in *MEPoetLitCorp* and some of its textual variants, which offers some insight into their use. Two of the uses in *MS*, *by reason and right law* (lxxv) and *by reason and rightful domes* (lxxvi), most clearly relate to the meaning attached to the original legal binomial in referring to the need to rule fairly, and with the guidance of rational judgement. There are also indirect references to this legal meaning in the construct *against reason and right* (lxx-lxxi). Their use broadly refers to reason as a general concept, but it would be unlikely that there would be no residual legal significance in the grammatical and lexical coordination of these two concepts. This indirect legal reference can also be ostensibly applied to statements beginning with *(h)it is reason and right*, before moving on to the conviction held by that particular individual (lxxix-lxxx).

Figure 6.19: Concordance of the 13 instances of ‘right’ and ‘reason’ in a binomial construction in the annotated *MEPoetLitCorp*, arranged alphabetically by L1 position.

(lxx)	hym, if thou might, Both again	**reason**	and right. I will no more in t
(lxxi)	eat, was grauntid vlixes Aganys	**reason**	and right, as the Renke thugh
(lxxii)	king that ye knew ever: For, as	**reason**	and rith rehersid to me ones,
(lxxiii)	eid. Thair is na ridand roy, be	**reason**	and right, Sa deir welcum this
(lxxiv)	tide in sorrow; He rulet him by R	**eason**	& the right spake. Songis of s
(lxxv)	ire lowe hertz To rule thaym by	**reason**	and by right law. Thay should
(lxxvi)	in all works, To rule thaym by	**reason**	and by rightful domes, Thorou
(lxxvii)	worthily wight - He shall have	**reason**	and right! Whethen is the coml
(lxxviii)	comes an errant knight. Do him	**reason**	and right For thy manhede."
(lxxix)	fallyn to the best. Syn hit is	**Reason**	& right, that renkes so many,
(lxxx)	, and fraite of our way. Hit is	**reason**	and right for your Rank love,
(lxxxi)	we reck it but little; It is	**reason**	and right the man be restrai
(lxxxii)	s emange duellis, For always on	**reason**	and on ri?t rewelle we oure-se

It is worth observing that all the rearranged binomials listed above in Figure 6.19 only appear in texts within *ARCorp* (including *aMA, AoA, GG, RtR, DT, MS,* and *WA*). Their absence in other alliterative texts, most notably the *Pearl* poems, is perhaps more indicative of differences in content as opposed to any significant changes in the underlying patterns of alliteration. While the presence of these binomials in so many of the alliterative texts is unsurprising, given the stylistic necessity for many ready-made alliterating linguistic units in accordance with the poetic demands of its metrical structure, the ability for these binomials to still convey legal meaning despite its syntactic reordering calls the formulaic nature of binomials into question. This is a topic which has recently been addressed by Mollin (2014), and would be interesting to integrate in a wider analysis of the (re)presentation of legal binomials and their rearranged forms within the reference corpus.

Given the comparative data on the use of *reason* between the two corpora, it appears that the primary stylistic difference setting *Piers Plowman* apart from the rest of the reference corpus – at least in this specific case – is the heavier use of exact legal phrases in preference to more creative versions of this recognisably legal language.

While there are differences in the frequencies of usage between the different identifiable meanings and functions of *reason* across the corpora, the range exhibited by these meanings remains unchanged; *PPCorp* does not contain any uses of *reason* unseen in *MEPoetLitCorp*, and vice versa. It also provides some insight into the formation of allegory in these poetic texts, and the ways in which they draw on various senses of *reason* in the formation of characters within *PP* and *RotR*, respectively. However, assessment of the larger legal-linguistic landscape across these medieval poems will have to wait for a more comprehensive study to ascertain whether the behaviour demonstrated by *reason* is representative of a larger pattern demonstrated by other KWs, or whether it exists as an outlier from conventional poetic use of legal language.

6.5 Conclusion

The present chapter used a corpus approach to address the legal themes, and the key lexis representing this theme, within *PP*. Initial KW analysis identified the text's statistically significant legal terms compared to the other poems in *MEPoetLitCorp*, while collocational analysis revealed the recurring linguistic mechanisms contributing to specific effects in the representation of the law (e.g., the stylistic use of *doctor* and *justice* and the binomial creativity evidenced in semi-techical terms like *reason*). This analytical focus on *doctor, justice,* and *reason* as bridging or 'split' words in Mahlberg and McIntyre's (2011) framework enabled this chapter to be oriented around those KWs demonstrating the most ambiguity in their textual functionality, and thus examples of words likely to be of high stylistic potential in their use.

Examining *doctor* and *justice* using Hoey's (2005) theory of lexical priming in 6.3 identified the animacy-based linguistic methods through which additional layers of pragmatic metaphorical significance are attributed to instances of class allegory in the poem. These techniques (collocation with markers of definiteness) were shown not to be unique to *PP* which suggests the presence of less overt allegorical constructs within medieval literary text-types in general, even those not generally attributed to the literary genre of allegory. This analysis additionally built on the theories and analysis of property allegory developed in Chapter 5, and moreover offered insight to terms which were originally unclassified in Mahlberg & McIntyre's (2011) framework when addressing the text's consumption-based KWs (e.g, *brewer*).

The analysis of *reason* in 6.4 to some degree contributed to the previous study of personification allegory but was further illustrative in demonstrating how technicality is a fundamental element embedded in the multiple thematic signalling and world-building functions ascribed to the KW in the poem, particularly in collocational patterns involving sub technical terminology. The prototypical range of pragmatic meanings attributed to reason in *PP* provided a point of departure for exploring the comparative use of the term within the general reference corpus, with the technicality evidenced in *PP* ultimately shown to be higher than the rest of the poetic texts in *MEPoetLitCorp*; however, the observed creativity evidenced in the literary reversal of word order within real-world legal binomials was shown to operate as part of a larger poetic tradition as a register- rather than stylistic-based feature.

7. CONCLUSION

This concluding chapter reflects on the analytical and methodological underpinnings of the present study while considering the implications of the study's findings. 7.1 starts by summarising the study's main findings as revealed through corpus analysis, and points out what these results have revealed about the stylistic patterns within (and beyond) *PP*. 7.2 then reflects on some of the leading methodological challenges and analytical limitations evidenced in this project, particularly concerning some of the impacts of corpus annotation originally explored in Chapter 3. The final section of this chapter outlines possible areas of future research in using quantitative linguistic tools and methods in the further study of language patterns in Middle English literature.

7.1 Main findings

This study is the first of its kind to examine the style of Middle English literary texts using corpus-assisted analysis. Exemplifying one of the most chronologically historical literary corpus stylistic studies to date, it pushes corpus stylistics as a field beyond its present focus on the style of Shakespeare and his literary contemporaries to the rich poetic tradition of Ricardian poetry. This has been conducted by looking at the key lexical patterns and terminology characterising *PP* compared to a sample reference corpus of contemporary poetic literature, with chapters addressing the prominent – and, at times, overlapping – themes of law and consumption generated from the corpus data. Each of these themes was explored using demarcated sets of key lexical terms through which to scrutinise the construction of ideologies surrounding the law and consumption and their presentation within the text's poetic and historical context. This language-oriented approach highlighted the recurring and underlying patterns

shaping these thematic constructs, which was used to further investigate how these aligned with (or diverged from) the patterns observable in other contemporary literary works using corpora comprised of just under 900,000 words of collected narrative poetry.

One of the key aspects of this study has been demonstrating how quantitative and qualitative methodologies can be productively combined to meet the demands of historical corpus stylistic analysis. This mixed-methods approach has provided empirical frequency and keyword data to corroborate qualitative observations made about the text's language use, either in this study or in existing critical scholarship. This approach has also been shown to offer statistically-based approximations of relative normality within which to situate qualitative observations made about the usualness (or in the case of unusualness, distinctiveness) of particular stylistic constructions in the text. These approximate measures of normality offer a way of mitigating the large temporally-based cognitive distance between modern reader and the poem's medieval origins, particularly through the loss of intuitive recognition of atypicality in the use of historical linguistic constructs on which stylistic interpretations of foregrounding and deviation are heavily dependent.

The stylistic patterns identified for *PP* were both lexical (KWs) and lexico-grammatical (KW collocations). These observations went beyond comparative evaluation that '*X* word occurs with greater statistical frequency in *PP* compared to other contemporary works' and '*X* word consistently co-occurs with *Y* word' to reveal the overarching stylistic effects generated by these salient language patterns. For example, the higher amounts of ambiguity present in the poem's personification allegory (e.g. *hunger*, *doctor*) has been shown to be a stylistic element generated from the poem's (in)animacy patterns which, while present in *PP*, are not replicated (or if so,

on the same scale) in the reference corpus texts. This stylistic distinctiveness was also identified in the poem's overwhelmingly negative approach towards consumption (*drink*), which has been shown to be underpinned by collocational patterns using lexis that generates meanings associated with perceptions of *propriety*; this was evidenced to some degree in *MEPoetLitCorp* (particularly *ARCorp*) but the texts in the reference corpus were instead shown to be characterised by the combination of multiple evaluative dimensions simultaneously (for example, *propriety* +*normality/security*). These findings are necessarily set apart from results which, after detailed comparative analysis, were found to be markers of register across my literary text-types (i.e., the use of creatively-revised binomials stemming from general and sub-technical legal language shown with *reason*).

These findings have significance in how style addressed in a medieval critical context. First, that 'style' as a concept cannot be exclusively linked to language patterns and features but needs to be linked to the interpretative effect(s) produced by these textual patterns in able to meaningfully encompass the distinctions between texts and the comparative relevance of these differences. Secondly, assertions of stylistic differences need to be methodologically transparent and empirically-based so as to render these claims falsifiable, including those made in the present investigation. While current medieval critical studies have revealed a wealth of potentially interesting stylistic information about different literary texts, too many claims at present are issued by leading names in the field whose seemingly untouchable reputations and authority (along with the lack of clearly identified methods) may inadvertently result in a form of scholarly gatekeeping. Corpus data and methods offer a way of not only welcoming new scholars into existing areas of stylistic inquriy – from

both stylistics and medieval studies – but also the means for performing new types of comparative analytical study of these literary text-types.

From a linguistics perspective, this investigation confirms the temporal durability of both corpus linguistic and stylistic theory when applied to ME texts. Within the remit of corpus linguistics, my expectation of the highly significant KW results producing key lexis denoting the names of main characters and main thematic topics addressed by the text, for example, mirror the projections set out by McEnery and Wilson (2001) , Scott and Tribble (2006), and Culpeper and Demmen (2015). This continuity was also seen in the collocational analysis conducted for select KWs in the poem, which broadly aligned with the pre-analysis literature-informed expectations set out in studies such as McEnery and Wilson (2001), Louw (2000), and Xiao (2015). There were some minor differences that departed from expectation in both KW and collocational analysis, but in both cases these were largely based on the presence of residual spelling variation in the corpus data as opposed to being associated with any impact-bearing factors tied to the historical nature of the text or its poetic form.

This investigation has especially contributed to the further development of Mahlberg and Mcintyre's (2011) corpus stylistic framework for functionality in literary KWs. In their initial introduction of the classification system Mahlberg & McIntyre (2011: 209-10 identified the fact that, while most polysemous and metaphoric language would possess inherent literary value as thematic signals, they would most likely prove difficult to categorise. My focus on the personification allegory in *PP* within 5.3 and 6.3 confirmed this difficulty with lexis being used metaphorically, and additionally illustrated how the personification allegory in *PP* embodies a rich mixture of comparatively foregrounded levels of text functionality wherein property allegory is generally primed as a fictional world element and class allegory more generally

operating as a thematic signal. This distinction is illustrated in Figure 7.1 below, thus revealing the main contribution of the current investigation to this framework in suggesting a level of nuance in (and comparative distribution of) allegorical behaviour that would otherwise be lost in broadly classifying these constructs universally as thematic signals.

Figure 7.1: Mahlberg & McIntyre's (2011) functional categories applied to *PP* as a single text, drawing from the top 285 lexical key words with all personification allegory highlighted in bold.

Category	Example key words
Fictional world	
Characters:	
Names	***conscience,*** *Piers,* ***mede, dowel, holy*** *(****church****),* *Jesus,* ***kind*** *(****wit****),* ***poverty, truth****, Christ****, reason, dobet, mercy****,* ***patience,*** *Eve,* ***dobest, pees, charity, hunger, guile, grace****, Lucifer, Adam,* ***repentance, symonye, contrition****,* *Marie* 'Mary', ***lechery,*** *Abraham,* ***faith***
Collective referents	*community, people, mankind*
Settings and props:	
Places	*Malvern*
Religious persons and institutions	*clergy, church, friars, saints, lollards****,*** *bishops, mass, monks, Christendom*
Courts and the law	*clerks, witness(es), writ*
Food and drink	*food, drink, ale, beggars*
Thematic signals	
	pore 'poor', *soul, wicked, pardon, heaven, suffering,* ***friar****, labourers, hell,* ***doctor****, needy, pilgrims, salvation, work,* ***reave*** 'reeve', *darkness*

7.2 Methodological reflections and limitations

The pattern-driven corpus approach taken in the current study, while yielding useful insights into the style- and register-based lexicogrammatical patterns present in *PP*, has also presented a series of challenges and some limitations. Firstly, even as comparatively small specialised literary corpora (Mahlberg & Smith 2010, McEnery & Wilson 2001: 117), *PPCorp* and *MEPoetLitCorp* contain far more emergent lines of inquiry than was possible to examine in the present investigation. Because of this, these corpora still contain sizable amounts of data which, if explored further, is likely to reveal further dimensions of knowledge about the language characteristically used in these late medieval poetic texts.

Secondly, focus on KWs and their collocations as the corpus data driving the focus of my research meant that analysis rans the risk of overlooking potentially significant stylistic features manifested in less frequent or non-statistically salient linguistic patterns. Highly creative or stylised language features used in a single instance carry equal analytical significance when examining style, but will not appear in the frequency-driven results yielded from a corpus approach. Additionally, more fine-grained analysis was limited to six KWs out of the 855 generated for the poem; this suggests that further avenues of corpus exploration can be found not only outside but also within the generated corpus data if redirecting focus to other key patterns.

Beyond these limitations in overall analytical scope and directionality, most of the challenges evidenced in this study were based on the possibilities set by the pre-analysis annotation produced for *PPCorp* and *MEPoetLitCorp*. While a successful first foray into orthographically regularising Middle English, my annotated corpora did not

contain enough resulting normalisation to ensure reliable corpus data was produced for the poem's grammatical KWs. This was because the relationship between form and function in most grammatical words is context-dependent, meaning that regularisation was only effective in the 32,000 word manual traning stage (Archer 2015). This was a keenly felt limitation, as the patterns revealed by these highly key function words have the potential to reveal newer stylistic insights than presented by the current focus on content KWs, which predominantly provided evidentiary support to corroborate findings already being discussed in the poem's literary criticism. In this sense, the innovative nature of the current project can be more clearly associated with the novelty of the (corpus) method itself and the future analytical potential of data mining in literary analysis over the originality of the findings revealed in the method's application to the poem in this instance.

In addition to amounts of modernisation achieved in the *VARD*ing process, the estimated percentage of accuracy in the annotated output for both *PPCorp* and *MEPoetLitCorp* was relatively low compared to other studies regularising historical datasets. One reason for this was the 50% confidence threshold used in this study, which meant that a higher amount of error was possible for normalisations than a higher threshold like 80% or even 90% through lower precision scores (Baron & Rayson 2009: 15). The lack of time to comparatively evaluate these thresholds to reveal the most optimal setting for my ME data was a challenge, one that also contributed to the reason for the relatively minimal amounts of manual annotation conducted on the corpora compared to the number of words present in the corpora themselves (32,000 versus 961,319). While the 32,000 words annotated in the present corpus exceeds Baron & Rayson's (2009: 21) suggestion of 12,000 as a ballpark for manual annotation and was enough to provide viable corpora for KW and collocational searches,

conducting greater amounts of manual annotation would have doubtlessly resulted in the tool being trained on a larger amount of variant forms.

These annotation-based constraints on what the study could analytically achieve are situated alongside limitations based on corpus composition. The 28 texts comprising my reference corpus are not inclusively representative of late 14th/early 15th century literary traditions, and further examination of the language patterns present in these texts would benefit from a wider sampling of literature than was accessed in the current investigation. There were also practical limitations on the types of source-texts that could be acquired for this study; not all the manuscriptal versions of the chosen poetic works were publicly available in machine-readable formats but most could be found in critical editions that had been digitised. This meant that *MEPoetLitCorp* did not mirror *PPCorp* in containing manuscript-sourced language, which resulted in the differing composition of each corpus.

This survey of the methodological challenges, and analytical restrictions caused by them, is not meant to devalue the findings of this corpus investigation, nor undermine the value of corpus approaches to the study of medieval language patterns. Rather, this overview is intended as an acknowledgement of some of the shortcomings identified in the corpus-driven approach I have taken here, shortcomings that are not only necessary in order to contextualise the outcomes of this study but also for signposting useful considerations for future strands of research at the intersection of computational methods and language analysis. The next and final part of this chapter below considers some possibilities for future research into late ME literary style.

7.3 Future directions

The present study offers a snapshot view into the ways in which a corpus linguistic approach can contribute to and further develop critical literary analysis of medieval literature like *PP*, but does not itself represent the full analytical potential of corpus methods that can be applied to the text. In this way, my research is hopefully representative of the first of many corpus-assisted analyses of *PP* and other medieval poetic text-types, with future studies extending beyond – and further contributing to – what findings the current investigation has to offer.

One avenue of future research could look anaphora, which would constitute an interesting feature of analysis through which to further study instances of personification allegory. Given the primary remit of this study in focusing on the poem's lexical KWs, this area unfortunately fell outside of the scope of the current research. However, given the increased amount of referential ambiguity already embedded in the use of pronominal anaphora, it would be interesting to see whether the same types of ambiguity were taking place, or whether any new linguistic patterns emerged in these contexts not seen in the direct reference to these personified entities by name (e.g. *he said* versus *Reason said*). This serves as a strand of further research into personification allegory in ME literature alongside others, such as looking at other types beyond property and class allegory (for example, architectural allegory) for evidence of new patterns in linguistic behaviour.

Another branch of future research could address the lack of machine-readable ME source-texts currently available for corpus stylistic research. This research would be heavily invested in resource creation, ideally with these resulting plain-text files being made open-access for the benefit of all researchers wanting to take a corpus approach to these medieval literary texts. This type of research would be heavily time-

intensive but make a substantive difference in increasing the attractiveness of corpus methods; it also coexists along future research strands prioritising focus on non-canonical texts or those for whatever reason are traditionally overlooked by focus on literary bastions such as *PP*, *The Canterbury Tales*, and *Confessio Amantis*.

Finally, and perhaps most urgently, future corpus research into ME text-types needs to address how to improve orthographic annotation in Middle English datasets. This approach could experiment with different settings in *VARD* to improve the annotated output, including studies comparing different confidence thresholds and amounts of manual annotation. Such investigations could also evaluate the comparative benefits of the type of modernisation (as done in the current investigation) with regularisation to an identified Middle English regularised system (i.e., lemmatisation to a universally recognised orthographic norm such as the headwords in the MED). Another possible avenue of further examination could study different types of orthographic regularisation through edit-distance programs like *VARD* compared to programs built using, for example, neural networks. Facilitating improvements in orthographic regularisation can thus lead to greater possibilities in studying recurring lexical patterns at the stylistic and register levels through corpus features such as n-grams, clusters, and lexical bundles; this did not lie within the remit of the current investigation but would be particularly effective in large-scale examination of formulaic phrases found within the alliterative revival texts like *PP*.

REFERENCES

Adams, R. (1988). Langland's theology. In J. Alford (Eds.) *A companion to Piers Plowman.* Berkeley: University of California Press, pp. 87-116.

Aers, D. (1975). *Piers Plowman and Christian allegory*. London: Arnold.

Aers, D. (1980). *Chaucer, Langland, and the creative imagination*. London: Routledge & Kegan Paul.

Aers, D. (2015). *Beyond reformation: an essay on William Langland's Piers Plowman and the End of Constantine Christianity.* Notre Dame: University Press.

Alexander, M., Dallachy, F., Piao, S., Baron, A., & Rayson, P. (2015). Metaphor, popular science, and semantic tagging: distant reading with the Historical Thesaurus of English. *Digital Scholarship in the Humanities* (doi:10.1093/llc/fqv045).

Alford, J. (1977). Law and literature in Medieval England. *PMLA*, 92, 941-51.

Alford, J. (1988a). *Piers Plowman: glossary of legal diction*. Cambridge: Brewer.

Alford, J. (1988b). The idea of Reason in Piers Plowman. In D. Kennedy et.al (Eds.), *Medieval English studies presented to George Kane* (pp.199-215). Woodbridge: Boydell & Brewer.

Alford, J. (1988c) (Ed.). *A companion to Piers Plowman*. Berkeley: University of California Press.

Anderson, T. & Crossley, S. (2011). Rue with a difference: a computational stylistic analysis of the rhetoric of suicide in Hamlet. In M. Ravassat and J. Culpeper (Eds.) *Stylistics and Shakespeare's language: transdisciplinary approaches*. London: Continuum, pp. 192-214.

Anipa, K. (2012). The use of literary sources in historical sociolinguistic research. In J. Hernández-Campoy & J. Conde-Silvestre (Eds.), *The handbook of historical sociolinguistics*. London: Wiley Blackwell, pp.170-190.

Archer, D., Wilson, A, & Rayson, P. (2002). *Introduction to the USAS category system*. Available at: http://ucrel.lancs.ac.uk/usas/usas_guide.pdf (accessed July 2016).

Archer, D., McEnery, T., Rayson, P., & Hardie, A. (2003). Developing an automated semantic analysis system for Early Modern English. In D. Archer, P. Rayson, A. Wilson,

& T. McEnery. (Eds.), *Proceedings of the Corpus Linguistics 2003 conference* (pp.22-31). Lancaster: UCREL.

Archer, D., & Rayson, P. (2006). *Dealing with variation: spelling. SCOTS symposium on linguistic variation and electronic projects*. University of Glasgow.

Archer, D. & Bousfield, D. (2010). 'See better, Lear'? See better, Lear! A corpus-based pragmastylistic investigation of Shakespeare's King Lear. In D. McIntyre & B. Busse (Eds.) *Language and style*. Basingstoke: Palgrave, pp. 183-203.

Archer, D., Kyto, M., Baron, A., & Rayson, P. (2015). *Guidelines for normalising Early Modern English corpora*. ICAME Journal, 39, 5-24.

Auer, A., Peersman, C., Pickl, S., Rutten, Gijsbert, & Vosters, R. (2015). Historical sociolinguistics: the field and its future. *Journal of Hisorical Sociolinguistics* 1(1), pp.1-13.

Baker, J. (2002). *An introduction to English legal history*. London: Reed Elsevier.

Baker, P. (2006). *Using corpora in discourse analysis*. London: Continuum

Baker, P. (2011). Times may change but we will always have money: diachronic variation in recent British English. *Journal of English Linguistics* 39(1), 65-88.

Baker, P., Hardie, A., & McEnery, T. (2006). *A glossary of corpus linguistics*. Edinburgh: University Press.

Baldwin, A. (1981). *The theme of government in Piers Plowman*. Cambridge: University Press.

Barnes, R. (1995). Langland's stance and style. *Yearbook of Langland Studies* (9), pp.19-37.

Barney, S. (2006). *The Penn commentary on Piers Plowman, volume 5*. Pennsylvania: University of Penn Press.

Baron, A. (2008). *Variant Detector* (Version 2.5.4). Lancaster University.

Baron, A. & Rayson, P. (2008). VARD2: a tool for dealing with the spelling variation in historical corpora. In *Proceedings of the Postgraduate Conference in Corpus Linguistics*. Birmingham: Aston University.

Baron, A., & Rayson, P. (2009). Automatic standardisation of texts containing spelling variation. How much training data do you need? In M. Mahlberg, V. Gonzalez-Diaz, & C. Smith (Eds.) *Proceedings of the Corpus Linguistics 2009 conference*, Liverpool.

Baron, A., Rayson, P., & Archer, D. (2009a). Automatic standardisation of spelling for historical text mining. In *Proceedings of Digital Humanities* 2009, Maryland, pp.309-312. College Park MD: University of Maryland.

Baron, A., Rayson, P., & Archer, D. (2009b). Word frequency and keyword statistics in historical corpus linguistics. *Anglistik: International Journal of English Studies* 20(1), 41-67.

Baron, A., Rayson, P., & Archer, D. (2011). *Quantifying Early Modern English spelling variation: change over time and genre.* Conference on New Methods in Historical Corpora, University of Manchester.

Barr, H. (1993). *The Piers Plowman tradition*. London: J.M Dent.

Barr, H. (1994). *Signes and sothe: language in the Piers Plowman tradition*. Cambridge: D.S. Brewer.

Barr, H. (2014). Major episodes and moments in Piers Plowman B. In A. Cole and A. Galloway (Eds.) *The Cambridge companion to Piers Plowman*. Cambridge: University Press, pp.15-32.

Barrington, C. (2010). John Gower's legal advocacy and 'in praise of peace'. In E. Dutton, J. Hines, & R. Yeager (Eds.), *John Gower, Trilingual Poet: language, translation, and tradition*, pp.112-125.

Bart P, Calabrese M, Duggan H, Duggan G and Turville-Petre T (Eds) (2018) *Piers Plowman: the text of the X manuscript.* Unpublished digital transcription of MS Hm 143 (accessed 26 August 2015).

Baugh, A. & Cable, T. (2002). *A history of the English language*. 5th Edition. London: Routledge.

Bennett, J. (1996). *Ale, beer, and brewsters in England: women's work in a changing world 1300-1600.* Oxford: University Press.

Benson, C. (2004). *Public Piers Plowman: modern scholarship and late medieval English culture*. Philadelphia: Penn State University Press.

Bhatia, V. K. (1987). Language of the law. *Language Teaching,* 20 (4), 227-234.

Bhatia, V.K. (1993). *Analysing genre: language use in professional settings*. London: Pearson Education Ltd.

Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing*, 8, 243-257.

Biber, D., Conrad, S., Finegan, E., Leech, G., & Johansson, S. (1999). *Longman grammar of spoken and written English.* Harlow: Longman.

Biber, D., Conrad, S., & Cortes, V. (2003). Lexical bundles in speech and writing: an initial taxonomy. In A. Wilson, P. Rayson, & T. McEnery (Eds.), *Corpus linguistics*. Frankfurt: Peter Lang.

Biber, D. & Conrad, S. (2009). *Register, genre, and style*. Cambridge: University Press.

Birnes, W. (1975). Christ as advocate: the legal metaphor of Piers Plowman. *Annuale Medievale*, 16 (71-93).

Blake, N. F. (1996). *A history of the English language*. Basingstoke: Macmillan.

Bloom, K., Garg, N., & Argamon, S. (2007). *Extracting appraisal expressions. Proceedings of the NAACL HLT 2007*.

Bloomfield, M. (1970). *Essays and Explorations: Studies in Ideas, Language and Literature.* Cambridge, MA: Harvard University Press.

Bondi, M (2010). Perspectives on keywords and keyness. In M. Bondi & M. Scott (Eds.), *Keyness in texts*. Amsterdam: John Benjamins, pp. 1-18.

Borroff, M. (1962). *Sir Gawain and the Green Knight: A Stylistic and Metrical Study*. New Haven: Yale University Press.

Bousfield, D. (2010). Never a truer word said in jest: a pragmastylistic analysis of impoliteness as banter in Henry IV, Part I. In M. Lambrou and P. Stockwell (Eds.), *Contemporary stylistics*. London: Bloomsbury, pp.209-220.

Bray, J. (2010). Writing presentation, the epistolary novel and free direct thought. In D. McIntyre and B. Busse (Eds.) *Language and style*. Basingstoke: Palgrave, pp.388-401.

Bray, J. (2014). A portrait of historical stylistics. In P. Stockwell & S. Whiteley (Eds.), *The Cambridge handbook of stylistics* (pp. 485-500). Cambridge: University Press.

Brewer, C. (1996). *Editing Piers Plowman: the evolution of the text.* Cambridge: University Press.

Brewer, C., & Rigg, A. (1983) *Piers Plowman: a facsimile of the Z-Text in Bodleian Library, Oxford, MS Bodley 851.* Toronto: Pontifical Institute of Mediaeval Studies.

Briggs, J., & Monaco, R. (1974). *Metaphor: the logic of poetry*. New York: Pace University Press.

Britnell, R. (2000). Urban demand in the British economy, 1300-1600. In J. Galloway (Ed.) In: *Trade, Urban Hinterlands and Market Integration* c.1300-1600. CMH Working Paper Series.

Brouwer, E. (2010). Matching cognitive and aesthetic accounts of poetic interpretation. In G. Low, Z. Todd, A. Deignan & L. Cameron (Eds.), *Researching and applying metaphor in the real world* (pp. 245-264). Amsterdam: John Benjamins.

Brown, A. (1989). *The governance of late medieval England 1272-1461*. Stanford: University Press.

Brown, P. (2009). *A companion to Medieval English literature and culture, 1350-1500.* Oxford: Blackwell.

Burke, M. (2014). *The Routledge handbook of stylistics*. Routledge: London.

Burke, M. (2014). Rhetoric and poetics: the classical heritage of stylistics. In M. Burke (Ed.), *The Routledge handbook of stylistics*. Routledge: London, pp.11-30.

Burrow, J. A. (1971). *Ricardian poetry: Chaucer, Gower, Langland, and the Gawain poet.* London: Routledge.

Burrow, J. A. (2008). *Medieval writers and their work: Middle English literature 1100-1500.* Oxford: University Press.

Burrows, J. F. (2002). Delta: a measure of stylistic difference and a guide to likely authorship. *Literary and Linguistic Computing*, 17, 267-287.

Burrows, J. F. (2005). Who wrote Shamela? Verifying the authorship of a parodic text. *Literary and Linguistic Computing*, 22, 24-27.

Busse, B. (2007). The stylistics of drama: an Early Modern English example. In P. Stockwell and M. Lambrou (Eds.) *Contemporary Stylistics*. London: Continuum, pp.232-43.

Busse, B (2010a). The stylistics of drama: the reign of King Edward III. In M. Lambrou and P. Stockwell (Eds.), *Contemporary stylistics*. London: Bloomsbury, pp.232-243.

Busse, B. (2010b). Recent trends in New Historical Stylistics. In B. Busse & D. McIntyre (Eds.), *Language and style* (pp. 32-54). Basingstoke: Palgrave Macmillan.

Busse, B. (2014). (New) historical stylistics. In M. Burke (Ed.), *The Routledge handbook of stylistics* (pp.101-117). London: Routledge.

Carrell, H. (2006). Food, drink and public order in the London Liber Albus. *Urban History* 33(2), 176-94.

Carter, R. (2004). *Language and creativity: the art of common talk.* London: Routledge.

Carter, R. (2010). Methodologies for stylistic analysis: practice and pedagogies In B. Busse & D. McIntyre (Eds.), *Language and style* (pp. 55-68). Basingstoke: Palgrave Macmillan.

Carter, R. (2012). Coda: some rubber bullet points. *Language and Literature*. 21 (1), pp. 106-114.

Carter, R. (2014). Stylistics as applied linguistics. In P. Stockwell & S. Whiteley (Eds.), *The Cambridge handbook of stylistics* (pp. 77-86). Cambridge: University Press.

Chambers, R.W & Grattan, J.H.G (1909). Texts and documents: The text of Piers Plowman 1, the A-text. *Modern Language Review* 4(3), pp. 357-389.

Chambers, R.W. (1932). The continuity of English prose from Alfred to More and his school. In N. Harpsfield (Ed.), *The life and death of Sir Thomas More*. London: Early English Text Society.

Chambers, R.W., Haselden, R.B, & Schulz, H.C. (1935). The manuscripts of Piers Plowman in the Huntington Library and their value for fixing the text of the poem. *The Huntington Library Bulletin* (8), pp. 1-27.

Channell, J. (2000). Corpus-based analysis of evaluative lexis. In S. Hunston & G. Thompson (Eds.), *Evaluation in text: authorial stance and the construction of discourse.* Oxford: University Press, pp.38-55.

Chism, C. (2002). *Alliterative Revivals*. Philadelphia: University of Pennsylvania Press.

Chung, T., & Nation, P. (2003). Technical vocabulary in specialised texts. *Reading in a foreign language* 15 (2), 103-116.

Chung, T., & Nation, P. (2004). Identifying technical vocabulary. *System* (32), 251-263.

Church, K. and Hanks, P. (1990) Word association norms, mutual information and lexicography. *Computational Linguistics* 16(1): 22–29.

Claridge, C. (2012). Of fox-sized mice and a thousand men: hyperbole in Old English. In C. Suhr & I. Taavitsainen (Eds.), *Studies in Variation, Contacts, and Change in English* 11 (E-series). Accessed 02 August 2017.

Collins, M. (forthcoming). *The hoole book: a literary-linguistic study of cohesion in Thomas Malory's Morte Darthur.* Unpublished PhD dissertation, University of Birmingham.

Comrie, B (1989). *Language Universals and Linguistic Typology*. Oxford: Basil Blackwell.

Cook, A. (2011). Cognitive interplay: how blending theory and cognitive science reread Shakespeare. In M. Ravassat and J.Culpeper (Eds.), *Stylistics and Shakespeare's language.* London: Continuum, pp.246-268.

Cordell, J. (2014). *Historical corpus stylistics: lexical characterization and the Pearl poems*. Unpublished Masters dissertation, University of Nottingham.

Cordell, J. & Wilkins, K. (2017). Crossing academic boundaries: linking literary linguistics to Old English pedagogy. *Paper presented at the annual Teachers of Old English in Britain and Ireland conference*. University College Cork (20 October).

Cordell, J. (2018). Priming text function in personification allegory: a corpus-based approach. *Language and Literature* 27(3), 218-39.

Cortes Rodriguez, L. (2006). Los elementos de final de serie enumerativa del tipo y todo eso, o cosas asi, y tal, etc. Boletin de Linguistica 18 (26), 102-129.

Cowan (1974). Lexical and syntactic research for the design of EFL. *TESOL Quarterly* (8), 389-399.

Craig, D. H., & Kinney, A. F. (2009). *Shakespeare, computers, and the mystery of authorship*. Cambridge: University Press.

Crassons, K. (2010). *The claims of poverty: literature, culture, and ideology in late medieval England.* Notre Dame: University Press.

Crisp, P. (2001). Allegory: conceptual metaphor in history. *Language and Literature* 10(1): 5-19.

Crystal, D. (1970). New perspectives for language study: stylistics. *English Language Teaching*, 24, 99-106.

Crystal, D. (1987). *The Cambridge encyclopedia of language.* Cambridge: University Press.

Crystal, D. (2003). *The Cambridge encyclopedia of the English language*. Cambridge: University Press.

Culpeper, J. (2001). *Language and characterisation: people in plays and other texts*. Harlow: Longman.

Culpeper, J. (2002). Computers, language and characterisation: an analysis of six characters in Romeo and Juliet. In U. Melander-Marttala, C. Ostman, & M. Kyto (Eds.), *Conversation in Life and in Literature: Papers from the ASLA Symposium.* (Vol. 15, pp. 11-30). Uppsala: Association Suedoise de Linguistique Appliquee.

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), 29-59.

Culpeper, J. (2014). Developing keyness and characterisation: annotation. In D. Hoover, J. Culpeper, & K. O'Halloran (Eds.), *Digital literary studies: corpus approaches to poetry, prose, and drama.* Abingdon: Routledge, pp. 9-34.

Culpeper, J., & Kyto, M. (2010). *Early modern English dialogues: spoken interaction as writing*. Cambridge: University Press.

Culpeper, J., & Demmen, J. (2015). Keywords. In D. Biber and R. Reppen (Eds.), *The Cambridge handbook of corpus linguistics*. Cambridge: University Press, pp. 90-105.

Damova, P. (2007). The language of law: a stylistic analysis with a focus on lexical (binomial) expressions. (Unpublished doctoral dissertation). Masaryk University, Brno.

Danet, B., & Bogoch, B. (1992). From oral ceremony to written document: the transitional language of Anglo Saxon wills. *Language and Communication,* 12(2), 95-122.

Davlin, M.C. (1995). A response to R. Barnes' 'Langland's stance and style'. *Yearbook of Langland Studies* 9, pp.31-33.

Davlin, M.C. (2018). Style and stylistics in the work of Charles Muscantine: a retrospect. *The Chaucer Review* 53 (1), pp. 112-118.

Demmen, J. (2012). *A corpus stylistic investigation of the language style of Shakespeare's plays in the context of other contemporary plays*. Unpublished PhD dissertation, University of Lancaster.

Diaz-Vera, J. (2014). From cognitive linguistics to historical sociolinguistics: the evolution of Old English expressions of shame and guilt. *Cognitive linguistic studies* 1(1), 55-83.

Diemer, S. (2012). Orthographic annotation of Middle English corpora. In J. Tyrkko, T. Nevalainen, M. Rissanen, & M. Kilpio (Eds.), *Outposts of historical corpus linguistics: from the Helsinki corpus to a proliferation of resources*. Helsinki: VARIENG.

Donner, M. (1987). Agent nouns in Piers Plowman. *The Chaucer Review* 21(3): 374-382.

Du Boulay, F. (1991). *The England of Piers Plowman: William Langland and his vision of the 14th century.* Cambridge: D.S. Brewer.

Duggan, H. (1986). *The shape of the B-Verse in Middle English alliterative poetry*. Speculum, 61(3), 564-592.

Duggan, H. (1990). Langland's dialect and final –e. *Studies in the Age of Chaucer* 12, 157-191.

Duggan, H. (2007). Metre, stanza, vocabulary, dialect. In Brewer, D. and Gibson, J. (Eds*.), A companion to the Gawain-Poet*. Cambridge: Boydell and Brewer, 240-2.

Duggan, H. (2009). Notes on the metre of Piers Plowman: twenty years on. In J. Jefferson & A. Putter (Eds.), *Approaches to the metres of alliterative verse* (pp. 159-186). Leeds: University Press.

Dyer, C. (1983). English diet in the later middle ages. In T.H. Aston, P.R. Coss, C. Dyer, & J. Thirsk (Eds.), *Social relations and ideas: essays in honour of R.H. Hilton*. Cambridge: University Press, 191-216.

Dyer, C. (1988). Changes in diet in the Late Middle Ages: the case of harvest workers. *The Agricultural History Review* 36(1), pp.21-37.

Dyer, C. (1998). Did the peasants really starve in medieval England? In M. Carlin (Ed.), *Food and eating in medieval Europe*. London: Hambledon Press, 53-71.

Dyer, C. (2000). *Everyday life in medieval England*. London: Hambledon Press.

Economou, G. (1996). *William Langland's Piers Plowman: The C-version.* Philadelphia: University of Pennsylvania Press.

Evans, M. (2018). *Messy signals: Communication and interpretation in a traditional/digital editing project*. Paper presented at the Computational Methods for Literary-Historical Textual Scholarship conference. De Montfort University: 3-5 July.

Fabizak, M., & Hebda, A. (2010). Cognitive historical approaches to emotions: pride. In M. Winters, H. Tissari, & K. Allan (Eds.), *Historical cognitive linguistics*. Berlin: DeGruyter, 261-297.

Fauconnier G and Turner M (2002) *The way we think: conceptual blending and the mind's hidden complexities*. New York: Basic Books.

Fialho, O., & Zyngier, S. (2014). Quantitative methodological approaches to stylistics. In M. Burke (Ed.), *The Routledge handbook of stylistics* (pp. 329-345). Abingdon: Routledge.

Firth (1968). A synopsis of linguistic theory 1930–1955. In F. R. Palmer (Ed.), *Selected papers of J. R. Firth 1952–59, 1–32*. Bloomington: Indiana University Press.

Fischer-Starke, B. (2010). *Corpus linguistics in literary analysis*. London: Continuum.

Fisher, J. (1977). Chancery English and the emergence of standard written English in the fifteenth century. *Speculum* 52(4), 870-889.

Fitzmaurice, S. (2010). Literary discourse. In A. Jucker & I. Taavitsainen (Eds.), *Historical pragmatics* (pp.679-704). Berlin: Mouton de Gruyter.

Flowerdew (2001). Concordancing as a tool in corpus design. In M. Ghadessy, A. Henry, & R. Roseberry (Eds.), *Small corpus studies and ELT: theory and practice* (pp. 71-92).

Foley, W. & Van Valin, R. (1985). Information packaging in the clause. *Language Typology and Syntactic Description* 1: 282-364.

Fowler, R. (1981). *Literature as social discourse.* Batsford: Indiana University Press.

Galloway, A. (2001). Piers Plowman and the subject of law. *Yearbook of Langland Studies* 15, 117-40.

Galloway, A. (2006). *The Penn commentary on Piers Plowman, volume 1*. Pennsylvania: University of Penn Press.

Gasse, R. (2016). Hunger and the clerical canine: the dog as metaphor in Piers Plowman B. *Enarratio: Publications of the Medieval Association of the West* 20, 67-91.

Gavins J (2007) *Text world theory: an introduction*. Edinburgh: University Press.

Gibbons, J. (Ed.) (1994). *Language and the law*. Harlow: Longman.

Giovanelli, M. (2013). *Text world theory and Keat's poetry: the cognitive poetics of desire, dreams, and nightmares.* London: Bloomsbury.

Gonzalez-Diaz, V. (2014). I quite detest the man: degree adverbs, female language and Jane Austen. *Language and Literature* 23(4), pp.310-330.

Gorlach, M. (1991). *Introduction to Early Modern English*. Cambridge: University Press.

Green, R. (1999). *A crisis of truth: literature and law in Ricardian England*. Philadelphia: University of Pennsylvania Press.

Green, K. (2008). Deixis and the poetic persona.In R. Carter and P. Stockwell (Eds.) *The language and literature reader.* Abingdon: Routledge, pp.127-136.

Griffiths, L. (1985). *Personification in Piers Plowman*. Cambridge: Brewer.

Ha, A. & Hyland, K. (2017). What is technicality? A Technicality Analysis Model for EAP vocabulary. *Journal of English for Academic Purposes* 28, 35-49.

Hall, J. (1949). *Living law in a democratic society*. Indiana: Bobbs Merrill.

Halliday, M., & Matthiessen, M. (2004). *Introduction to Functional Grammar.* Abingdon: Routledge.

Hamilton, C. (2014). Stylistics as rhetoric. In P. Stockwell and S. Whiteley (Eds.), *The Cambridge handbook of stylistics.* Cambridge: University Press, pp. 63-76.

Hanna, R. (2006). Brewing trouble: on literature, history, and alewives. In B. Hanawalt & D. Wallace (Eds.), *Bodies and disciplines: intersections of literature and history in fifteenth-century England.* Minneapolis: University of Minnesota Press, pp.1-17.

Hanna, R. (2017). *The Penn commentary on Piers Plowman, volume 2*. Pennsylvania: University of Penn Press.

Harbus, A. (2006). *Cognitive approaches to Old English poetry.* Cambridge: D.S. Brewer.

Harbus, A. (2016). Autofocus and remote text-world building in the earliest English narrative poetry. In J. Gavins & E. Lahey (Eds.) *World building: discourse in the mind*. Switzerland: Palgrave MacMillan, pp. 545-576.

Harvey, B. (1977). *Westminster Abbey and its estates in the middle ages.* Oxford: Clarendon Press.

Harvey, K. (2013). *Investigating adolescent health communication*. London: Bloomsbury.

Hiltunen, T. & Tyrkko, J. (2013). Tagging Early Modern English Medical Texts. *Corpus Analysis with Noise in the Signal (CANS) 2013 workshop, Corpus Linguistics 2013*. Lancaster University.

Ho, Y. (2011). *Corpus stylistics in principle and practice: a stylistic exploration of John Fowles' The Magus*. London: Bloomsbury.

Hockey, S. (2000). *Electronic texts in the humanities*. Oxford: University Press.

Hoey, M. (2004) The textual priming of lexis. In: Aston G, Bernardini S and Stewart D (Eds) *Corpora and Language Learners.* Amsterdam: John Benjamins, pp.21-44.

Hoey, M. (2005) *Lexical Priming: A New Theory of Words and Language.* London: Routledge.

Holdsworth, W. (1982). *A history of English law, volume 1*. London: Methuen & Co. Ltd.

Hoover, D. (2004). Testing Burrows' Delta. *Literary and Linguistic Computing*, 19, 453-475.

Horobin, S. (2005). 'In London and opelond': the dialect and circulation of the C version of Piers Plowman'. *Medium Aevum*, 74, 248-269.

Hoover, D. (2007). Corpus stylistics, stylomtery, and the style of Henry James. *Style* 41(2), 174-203.

Horobin, S., & Smith, J. (2002). *An introduction to Middle English.* Edinburgh: University Press.

Horobin, S. (2003). *The Language of the Chaucer Tradition*. Cambridge: D.S. Brewer.

Horobin, S. (2012). *Chaucer's Language*. 2nd edition. Basingstoke: Palgrave Macmillan.

Horobin (2013). *Does spelling matter?* Oxford: University Press.

Horton, T. (1987). *The effectiveness of the stylometry of function words in discriminating between Shakespeare and Fletcher*. Unpublished PhD dissertation, University of Edinburgh.

Hughes, J. (1995). A response to R. Barnes' 'Langland's stance and style'. *Yearbook of Langland Studies* 9, pp. 33-7.

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: University Press.

Hunston, S. (2011). *Corpus approaches to evaluation: phraseology and evaluative language*. London: Routledge.

Hunston, S. & Thompson, G. (2000). Evaluation: an introduction. In S. Hunston & G. Thompson (Eds.), *Evaluation in text: authorial stance and the construction of discourse.* Oxford: University Press, pp.1-27.

Inman, M. (1978). Lexical analysis of scientific and technical prose. In L. Trimble, M. Todd-Trimble, & K. Drobnic (Eds.), *ESP: Science and technology,* pp. 242-256.

Jaafar E (2017) Corpus stylistic analysis of Thomas Harris' The Silence of the Lambs. Khazar Journal of Humanities and Social Sciences 20(1): 25-42.

Jeffries, L. & McIntyre, D. (2010). *Stylistics.* Cambridge: University Press.

Jensen, K. (2014). Linguistics in the digital humanities: (computational) corpus linguistics. *Journal of media and communication research*, 30, 20. Online. Retrieved 06 August 2015.

Jenset, G. & McGillivray, B. (2017). *Quantitative historical linguistics: a corpus framework.* Oxford: University Press.

Jones, M. (2011). *Radical pastoral, 1381-1594: appropriation and the writing of religious controversy.* Farnham: Ashgate.

Kane, G. (1960). *Piers Plowman, the A version: an edition in the form of Trinity College Cambridge MS R.3.14 corrected and restored from the known evidence, with variant readings.* London: The Athlone Press.

Kane, G. (1965). *Piers Plowman: the evidence for authorship*. London: Athlone Press.

Kane, G., and Donaldson, E. (1975). *Piers Plowman, the B version: an edition in the form of Trinity College Cambridge MS B.15.17 corrected and restored from the known evidence, with variant readings.* London: Athlone Press.

Kennedy, G. (1998). *An introduction to corpus linguistics*. London: Longman.

Kilgarriff, A. (2001). Comparing corpora. *International Journal of Corpus Linguistics* 6(1): 97-133.

Kim, M. (2004). The politics of consuming worldly goods: negotiating Christian discipline and feudal power in Piers Plowman. *Traditio* 59, 339-368.

Kirk, E. & Anderson, J. (1990). Introduction to *Piers Plowman: an alliterative verse translation by E.T. Donaldson*. New York: W.W. Norton & Company, pp.vii-xix.

Kirk, R. (1933). References to the law in *Piers the Plowman*. *Publications of the Modern Language Association* 48, 322-27.

Kopaczyk, J. (2013). *The legal language of Scottish Burgs: standardisation and lexical bundles 1350-1560*. Oxford: University Press.

Kopaczyk, J. & Tyrkko, J. (2018). *Applications of pattern-driven methods in corpus linguistics.* Amsterdam: John Benjamins.

Koskenniemi, I. (1968). *Repetitive word-pairs in Old and Early-Middle English prose.* Turku: Turun Yliopisto.

Lakoff, G., & Turner, M. (1989). *More than cool reason: a field guide to poetic metaphor.* Chicago: University Press.

Lakoff, G., & Johnson, M. (1980a). *Metaphors we live by*. Chicago: University Press.

Lakoff, G. & Johnson, M. (1980b). Conceptual metaphor in everyday language. *The Journal of Philosophy* 77(8), 453-86.

Lambrou, M. & Stockwell, P. (2010). *Contemporary stylistics*. London: Continuum.

Langacker, R. (1991). *Foundations of cognitive grammar: descriptive applications*. Stanford: University Press.

Lawlor, J. (1962). *Piers Plowman: an essay in criticism.* London: Edward Arnold.

Lawrence, A. (2018). *AntConc* (Version 3.5.6). Tokyo: Waseda University.

Lee, D. (2010). What corpora are available? In A. O'Keefe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 107-121). Abingdon: Routledge

Leech, G. (1969). *A Linguistic Guide to English Poetry*. London: Longman.

Leech, G. (1985). Stylistics. In T. A. van Dijk (Ed.), *Discourse and literature* (pp. 39-57). Amsterdam: John Benjamins.

Leech, G. (2008). *Language in literature: style and foregrounding.* Harlow: Pearson Education Ltd.

Leech, G. (2010). Analysing literature through language: two Shakespearian speeches. In D. McIntyre and B. Busse (Eds.), *Language and style*. Basingstoke: Palgrave, pp.15-31.

Leech, G., & Short, M. (1981). *Style in fiction*. Harlow: Pearson Education Ltd.

Leech, G., Rayson, P., & Wilson, A. (2001). *Word frequencies in written and spoken English: based on the British National Corpus*. London: Longman.

Leech, G., & Short, M. (2007). *Style in fiction: a linguistic introduction to English fictional prose.* Harlow: Pearson Education Ltd.

Lehto, A., Baron, A., Ratia, M., & Rayson, P. (2010). Improving the precision of corpus methods: the standardized version of Early Modern English Medical Texts. In I. Taavitsainen & P. Pahta (Eds.), *Early Modern English Medical Texts: corpus description and studies* (pp. 279-289). Amsterdam: John Benjamins.

Levinson, S. C. (1979). Pragmatics and social deixis: Reclaiming the notion of conventional implicature. In C. Chiarello (Ed.), *Proceedings of the Fifth Annual Meeting of the Berkeley Linguistics Society* (pp. 206-223).

Levinson, S. C. (1983). *Pragmatics*. Cambridge: University Press.

Louw, B. (1993). Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In M. Baker, G. Francis, and E. Tognini-Bonelli (Eds.), *Text and technology*, 157–176. Philadelphia and Amsterdam: John Benjamins

Louw, N. (2000). Contextual prosodic theory: Bringing semantic prosodies to life. In C. Heffer, H. Sauntson, and G. Fox (Eds.), *Words in context: A tribute to John Sinclair on his retirement.* University of Birmingham.

Louw, N. & Milojkovic, M. (2016). *Corpus stylistics as contextual prosodic theory and subtext*. Amsterdam: John Benjamins.

McCarthy, M., & Handford, M. (2004). *Invisible to us: a preliminary corpus-based study of spoken business English*. Amsterdam: John Benjamins.

McEnery, T., & Wilson, A. M. (2001). *Corpus linguistics.* Edinburgh: University Press.

McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: an advanced resource book.* London: Routledge.

McEnery, T., & Hardie, A. (2012). *Corpus linguistics: methods, theory, and practice.* Cambridge: University Press.

McIntyre, D. (2010). Dialogue and characterisation in Quentin Tarantino's Reservoir Dogs: a corpus stylistic analysis. In B. Busse and D. McIntyre (Eds.) *Language and style*. Basingstoke: Palgrave, pp. 162-182.

Mahlberg, M. (2007). Corpus stylistics: bridging the gap between linguistic and literary studies. In M. Hoey, M. Mahlberg, M. Stubbs & W. Teubert (Eds.), *Text, discourse and corpora: theory and analysis* (pp. 219-246). London: Continuum.

Mahlberg, M. (2009). Corpus stylistics and the Pickwickian watering pot. In P. Baker (Ed.) *Contemporary corpus linguistics*. London: Continuum, pp. 47-63.

Mahlberg, M. (2010). A corpus stylistic perspective on Dickens' Great Expectations. In M. Lambrou & P. Stockwell (Eds.) *Contemporary stylistics*. London: Continuum, pp.19-31.

Mahlberg, M. (2013). *Corpus stylistics and Dickens' Fiction.* London: Routledge.

Mahlberg, M. (2014). Corpus stylistics. In M. Burke (Ed.), *The Routledge Handbook of Stylistics.* London & New York: Routledge, pp. 378-392.

Mahlberg, M. (2016). Corpus stylistics. In V. Sotirova (Ed.), *The Bloomsbury companion to stylistics*. London: Bloomsbury, pp.139-156.

Mahlberg, M., & Smith, C. (2010). Corpus approaches to prose fiction: civility and body language in Pride and Prejudice. In B. Busse & D. McIntyre (Eds.), *Language and style* (pp. 449-467). Basingstoke: Palgrave Macmillan.

Mahlberg, M. & McIntyre, D. (2011). *A case for corpus stylistics: Ian Fleming's Casino Royal.* English Text Construction. 4(2), 204-277.

Mahlberg, M., & Smith, C. (2012). Dickens, the suspended quotation and the corpus. *Language and Literature*, 51-65.

Mahlberg, M., Smith, C., & Preston, S. (2013). Phrases in literary contexts: patterns and distributions of suspensions in Dickens's novels. *International Journal of Corpus Linguistics* 18(1), pp.35-56.

Mair, C. (2006). *Twentieth century English*. Cambridge: University Press.

Maley, Y. (1994). The language of the law. In J. Gibbons (Ed.), *Language and the law* (pp.11-50). Harlow: Longman.

Malkiel, Y. (1959). Studies in irreversible binomials. *Lingua* (8), 113-160.

Mann, J. (1979). Eating and drinking in Piers Plowman. *Essays and Studies* 32, 26-43.

Mann, J. (1994). The power of the alphabet: a reassessment of the relation between the A and B versions of Piers Plowman. *The Yearbook of Langland Studies*, 8, 21-50.

Marin, M. and Rhea, C. (2014). Researching legal terminology: a corpus-based proposal for the analysis of sub-technical legal terms. *ASp* (66), 61-82.

Marin Perez, M. (2014). A proposal to exploit legal term repertoires extracted automatically from a legal English corpus. *Miscelanea: a journal of English and American studies* (49), 53-72.

Marcus, M. (1996). The complexity and diversity of the words in Wright's English Dialect Dictionary. In M. Markus, Y. Iyeiri, & R. Heuberger (Eds.) *Middle and Modern English corpus linguistics: a mulit-dimensional approach.* Amsterdam: John Benjamins, pp.209-224.

Markus, M. (2000). Normalising the word forms in The Ayenbite of Inwyt. In I. Taavitsainen, T. Nevalainen, P. Pahta, & M. Rissanen (Eds.) *Placing Middle English in Context*. Berlin: Mouton De Gruyter, pp.181-198.

Markus, M. (2002). Towards an Analysis of Pragmatic and Stylistic Features in 15th and 17th Century English Letters. In P. Peters, P. Collins & A. Smith (Eds.). *New frontiers of corpus research. Papers from the Twenty-First International Conference on EnglishLanguage Research on Computerized Corpora*, Sydney 2000 (Language andComputers series 36). Amsterdam and New York: Rodopi, 2002, pp.179-198.

Markus, Manfred. (2002) - In: New Frontiers of Corpus Research p. 179-198

Martin, J.R. (1991). Nominalization in science and humanities: distilling knowledge and scaffolding text. In E. Ventola (Ed.), *Functional and systemic linguistics: approaches and uses.* Berlin: Mouton de Gruyter, pp. 307-338.

Martin, J. (2000). Beyond exchange: appraisal systems in English. In S. Hunston & G. Thompson (Eds.), *Evaluation in text: authorial stance and the construction of discourse.* Oxford: University Press, pp.142-75.

Martin, J., & White, P. (2005). *The language of evaluation: appraisal in English*. London: Palgrave.

Massimiliano, M. (2010). The poetics of disengagement: Jane Austen and echoic irony. *Language and Literature* (19)4, pp, 339-356.

Mastropierro, L. (2017). *Corpus stylistics in Heart of Darkness and its Italian translations*. London: Bloomsbury.

Mellinkoff, D. (1963). *The language of the law*. New York: Little, Brown, & Co.

Middle English Dictionary (2013). 1st ed. Ann Arbor: University of Michigan Digital Collections.

Middleton, A. (1988). Introduction: the critical heritage. In J. Alford (Ed.), *A companion to Piers Plowman*. Berkeley: University of California Press, pp.1-28.

Mollin, S. (2012). Revisiting binomial order in English: ordering constraints and reversibility. *English Language and Linguistics* 16(1), 83-103.

Mollin, S. (2014). *The (ir)reversibility of English binomials*. Amsterdam: John Benjamins.

Mooney, A. (2014). *Language and law*. Basingstoke: Palgrave Macmillan.

Moriya, Y. (2014). *Repetition, regularity, redundancy: norms and deviations of Middle English alliterative meter.* Tokyo: Hituzi Syobo Publishers.

Muldner, G., & Richards, M. (2006). Fast or feast: reconstructing diet in later medieval England by stable isotope analysis. *Journal of Archaeological Science* 32, 39-48.

Murphy, S. (2015). I will proclaim myself what I am: corpus stylistics and the language of Shakespeare's soliloquies. *Language and Literature* 24(4), pp.338-354.

Muscatine, C. (1963). *Chaucer and the French tradition: a study in style and meaning*. Berkeley: University of California Press.

Muscatine, C. (1972). *Poetry and crisis in the age of Chaucer*. Notre Dame: University Press.

Muscatine, C. (1999). *Medieval literature, style, and culture*. Columbia: University of South Carolina Press.

Nevalainen, T. (2006). *An introduction to Early Modern English.* Oxford: University Press.

Norgaard, N., Montoro, R., & Busse, B. (2010). *Key terms in stylistics*. London: Continuum

O'Halloran, K. (2007a). Corpus-assisted literary evaluation. *Corpora* 2(1), pp.33-63.

O'Halloran, K. (2007b). The subconscious in James Joyce's "Eveline": a corpus stylistic analysis that chews on the "Fish hook". *Language and Literature* 16 (3): 227–244.

Oakden, J. (1935). *Alliterative poetry in Middle English: a survey of the traditions*. Manchester: University Press.

Oncins-Martinez, J. (2011). Shakespeare's sexual language and metaphor: a cognitive stylistic approach. In M. Ravassat and J.Culpeper (Eds.), *Stylistics and Shakespeare's language*. London: Continuum, pp.215-245.

Overstreet, M. (2015). The role of pragmatic function in the grammaticalization of English general extenders. Pragmatics: the quarterly publication of the International Pragmatics Association, 24 (1), 105-129.

Oxford English Dictionary (2016). 2nd ed. Oxford: University Press.

Pakkala-Weckstrom, M. (2010). Chaucer. In A. Jucker and I. Taavitsainen (Eds.), *Historical Pragmatics*. Berlin: De Gruyter Mouton, pp.219-246.

Pearsall, D. (1988). Book review (Alford 1988a). Piers Plowman: a glossary of legal diction. *Speculum* 65, pp.363-4.

Pearsall, D. (1990). *An annotated critical bibliography of Langland.* Hertfordshire: Harvester Wheatsheaf.

Pearsall, D. (Ed.). (1994). Piers Plowman, by William Langland: an edition of the C-text. Exeter: University Press.

Pearsall, D. (2006). Review. For A. Galloway (Ed.), *The Penn Commentary on Piers Plowman* (Vol. 1). Philadelphia: University of Pennsylvania Press.

Piao, S., Dallachy, F., Baron, A., Alexander, M. (2014). Developing the Historical Thesaurus Semantic Tagger. *Talk presented at the DHC conference*, September (Sheffield).

Pollock, F., & Maitland, F. (1895). *The history of English law before Edward I.* Cambridge: University Press.

Pons-Sanz, S. (2015). Old English style. In V. Sotirova (Ed.), *The Bloomsbury companion to stylistics*. London: Bloomsbury, pp.569-582.

Ratia, M. (2013). Performing keyword analysis on automatically vs. manually VARDed corpora: the case of Stuart plague treatises. *Corpus Analysis with Noise in the Signal (CANS) 2013 workshop, Corpus Linguistics 2013*. Lancaster University.

Ravassat, M. & Culpeper, J. (Eds.) (2011). *Stylistics and Shakespeare's language: transdisciplinary approaches.* London: Continuum.

Rayson, P. (2008). From keywords to key semantic domains. *International Journal of Corpus Linguistics* 13(4): 519-549.

Rayson, P. (2009) *Wmatrix: a web-based corpus processing environment,* Computing Department, Lancaster University. http://ucrel.lancs.ac.uk/wmatrix/.

Rayson, P., Archer, D., Baron, A., Culpeper, J., and Smith, N. (2007). Tagging the Bard: Evaluating the accuracy of a modern POS tagger on Early Modern English corpora. In *Proceedings of Corpus Linguistics 2007 conference*, Birmingham.

Rayson, P. & Baron, A. (2011). Automatic error tagging of spelling mistakes in learner corpora. In F. Meunier, S. De Cock, G. Gilquin, and M. Paquot (Eds.) *A Taste for Corpora: in honour of Sylviane Granger*, Studies in Corpus Linguistics, 45. John Benjamins, Amsterdam.

Rayson, P., Baron, A., Piao, S., & Wattam, S. (2015). Large-scale time-sensitive semantic analysis of historical corpora. *Talk presented at the ICAME conference*, Trier (May).

Rigg, G., and Brewer, C. (1983). *Piers Plowman: the Z version*. Toronto: University Press.

Rissanen, M. (1994). The Helsinki corpus of English texts. In M. Kyto, M. Rissanen, & S. Wright (Eds.), *Corpora across the centuries: proceedings of the First International Colloquium on English Diachronic Corpora*. Amsterdam: Rodopi.

Rissanen, M. & Tyrkko, J. (2013). The Helsinki Corpus of English Texts. In A. Meurman-Solin & J. Tyrkkö (Eds.), Principles and Practices for the Digital Editing and Annotation of Diachronic Data. Available at: http://www.helsinki.fi/varieng/series/volumes/14/rissanen_tyrkko/ (accessed April 2018).

Robbins, R.H. (1960). Middle English poems of protest. *Anglia* 78, pp.193-203.

Robertson, E. & Shepherd, S. (Eds.) (2006). *Piers Plowman, by William Langland: a Norton critical edition*. New York: W.W. Norton & Company.

Ruano San Segundo, P. (2016). A corpus stylistic approach to Dickens' use of speech verbs: beyond mere reporting. *Language and Literature* 25(2), pp.113-129.

Rudd, G. (1994). *Managing Language in Piers Plowman.* Cambridge: D.S. Brewer.

Russell, G., & Kane, G. (1997). *Piers Plowman, the C version: an edition in the form of Huntington Library MS HM 143, corrected and restored from the known evidence, with variant readings.* London: The Athlone Press.

Salter, E. & Pearsall, D. (Ed.) (1969). *Piers Plowman*. Evanston: York Medieval Texts.

Samuels, M. L. (1985). Langland's dialect. *Medium Aevum*, 54, 232-247.

Samuels, M. L. (1988). Dialect and grammar. In J. Alford (Ed.), *A companion to Piers Plowman*. Berkeley: University of California Press, pp. 201-221.

Savage, A. (1993). Piers Plowman: the translation of scripture and food for the soul. *English Studies* 74(3), 209-21.

Scanlon, L. (2007). *Personification and penance.* Yearbook of Langland Studies 21: 1–30.

Schane, S. (2006). *Language and the law*. New York: Continuum.

Schendl, H. (2000). Syntactic constraints on code-switching in medieval texts. In I. Taavitsainen, T. Nevalainen, P. Pahta, & M. Rissanen (Eds.) *Placing Middle English in context*. Berlin: Mouton de Gruyter: 67-95.

Schmidt, A.V.C. (1980). Langland's structural imagery. *Essays in Criticism* 30(4), 311-25.

Schmidt, A.V.C. (1987). *The clerkly maker: Langland's poetic art.* Cambridge: D.S. Brewer.

Schmidt, A.V.C. (1995a). *William Langland's Piers Plowman: A parallel-text of the A, B, C and Z versions* (Vol. 1: Text). London: Longman.

Schmidt, A.V.C. (1995b). *William Langland's Piers Plowman: A parallel-text of the A, B, C and Z versions* (Vol. 2: Introduction, textual notes, commentary, bibliography and indexical glossary). London: Longman.

Schmidt, A.V.C. (1995c). *The vision of Piers Plowman: the B text.* London: Everyman.

Schmidt, A.V.C (2011). *William Langland's Piers Plowman: A parallel-text of the A, B, C and Z versions* (Vol. 1: Text). Kalamazoo: Medieval Institute Publications.

Scott, M. (1997). PC analysis of keywords — and key keywords. *System*, 25 (2), 233–245.

Scott, M. (2000). Focusing on the text and its keywords. In L. Burnard & T. McEnery (Eds.), *Rethinking language pedagogy for a corpus perspective* (pp.103-122). Frankfurt: Peter Lang.

Scott, M. (2001). Comparing corpora and identifying keywords, collocations, and frequency distributions through the WordSmith Tools suite of computer programs. In M. Ghadessy (Ed.), *Small corpus studies and ELT: theory and practice*. Philadelphia: John Benjamins, 47-67.

Scott, M. (2012). *WordSmith Tools* (Version 6). Liverpool: Lexical Analysis Software.

Scott, M., & Tribble, C. (2006). *Textual patterns: keywords and corpus analysis in language education*. Amsterdam: John Benjamins.

Scragg, D.G. (1974). *A history of English spelling*. Manchester: University Press.

Semino, E., & Culpeper, J. (2011). *Stylistics.* In J. A. Ostman & J. Verschueren (Eds.), Pragmatics in practice (pp. 295-305). Amsterdam: John Benjamins.

Short, M. (1996). *Exploring the language of poems, plays and prose.* Harlow: Pearson Education Ltd.

Short, M. (2008). '"Where are you going to my pretty maid?" "For detailed analysis" sir, she said'. In G. Watson (Ed.), *The state of stylistics: PALA 26* (pp. 1-29). Amsterdam: Rodopi.

Simpson, J. (1990). *Piers Plowman: an introduction to the B-text*. Longman: New York.

Simpson, P. (2004). *Stylistics: a resource book for students*. Abingdon: Routledge.

Simpson, J. (2007). *Piers Plowman: an introduction*. Liverpool: University Press.

Simpson, P. (2014). *Stylistics: a resource book for students*. Second edition. Abingdon: Routledge.

Sinclair, J. (1966). Beginning the study of lexis. In C. Bazell, J. Catford, M. A. K. Halliday, and R. Robins (Eds.), *In memory of J. R. Firth*, 410-430. London: Longman.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: University Press.

Sinclair, J. (1992). The automatic analysis of text corpora. In J. Svartvik (Ed.), *Directions in corpus linguistics: proceedings of the Nobel symposium 82, Stockholm* (pp.379-397). The Hague: Mouton.

Sinclair, J. (1996). The search for units of meaning. *Textus,* 9, 75-106.

Sinclair, J. (1999). A way with common words. In H. Hasselgard & S. Oskefjell (Eds.), *Out of corpora: studies in honour of Stig Johnasson.* Amsterdam: Rodopi, pp. 157-79.

Sinclair, J. (2004). *Trust the text: language, corpus, and discourse*. London: Routledge.

Skeat, W. W. (1888). *The vision of Piers Plowman*. Oxford: Clarendon Press.

Smith, C., Adolphs, S., Harvey, K. & Mullany, L. (2014). Spelling errors and keywords in born-digital data: a case study using the Teenage Health Freak Corpus. *Corpora* 9(2), pp. 137-154.

Smith, J. (1988) (Ed.). *The English of Chaucer and his contemporaries*. Aberdeen: University Press.

Smith, J. & Samuel, M. (1981). The Language of Gower. *Neuphilologische Mitteilungen: Bulletin of the Modern Language Society*, 82, pp. 295-304.

Sotirova, V. (2015). *The Bloomsbury companion to stylistics*. London: Bloomsbury.

Spearing, A.C. (1970). *The Gawain poet: a critical study.* Cambridge: University Press.

Spitzer, L. (1948). *Linguistics and literary history.* Princeton: University Press.

Sprang, F. (2017). The confines of cognitive literary studies: the sonnet and a cognitive poetics of form. *Journal of Literary Theory* 11(2), pp.240-254.

Stanley, E.G. (2008). Allegory through the ages, as read mainly in England and as seen anywhere. In: M. Carr, K. Clarke, and M. Nievergelt (eds) *On Allegory: Some Medieval Aspects and Approaches.* Cambridge: Scholars Publishing, pp.1-27.

Steiner, E. (2003). *Documentary culture and the making of medieval English literature*. New York: Cambridge University Press.

Steiner, E., & Barrington, C. (2002). *The letter of the law: legal practice and literary production in medieval England*. Cornell: University Press.

Stockwell, P. (2002). *Cognitive poetics: an introduction.* London: Routledge.

Stockwell, P. (2005). Stylistics: language and literature. In B. Aarts & A. McMahon (Eds.), *The handbook of English linguistics* (pp. 742-758). Oxford: Blackwell.

Stockwell, P. (2009). The cognitive poetics of literary resonance. *Language and Cognition*, 1(1): 25-44.

Stockwell, P. (2017). *The language of surrealism*. London: Palgrave.

Stockwell, P & Whiteley (Eds.). *The Cambridge handbook of stylistics.* Cambridge: University Press.

Stockwell, P., & Whiteley, S. (2014).Introduction. In P.Stockwell and S.Whitely (Eds.), *The Cambridge handbook of stylistics* (pp.1-10). Cambridge: University Press.

Stockwell, P. & Mahlberg, M. (2015). Mind-modelling with corpus stylistics in David Copperfield. *Language and Literature* 24(2), pp. 129-147.

Stokes, M. (1984). *Justice and mercy in Piers Plowman*. London: Croom Helm.

Stubbs, M. (2005). Conrad in the computer: examples of quantitative stylistic methods. *Language and Literature*, 14(1), 5-24.

Studer, P. (2014). *Historical corpus stylistics: media, technology, and change*. London: Bloomsbury.

Sylvester, L. (2015). Middle English style. In V. Sotirova (Ed.), *The Bloomsbury companion to stylistics*. London: Bloomsbury, pp. 583-606.

Taavitsainen, I. (2001). Changing conventions of writing: the dynamics of genres, text-types, and text traditions *European Journal of English Studies,* 5(2), 139-150.

Taavitsainen, I., Nevalainen, T., Pahta, P., & Rissanen, M. (2000). Introduction. In I. Taavitsainen, T. Nevalainen, P. Pahta, & M. Rissanen (Eds.) *Placing Middle English in context*. Berlin: Mouton de Gruyter: 1-4.

Taavitsainen, I., Pahta, P., & Makinen, M. (2006). *Middle English Medical Texts*. Amsterdam: John Benjamins.

Taavitsainen, I., & Fitzmaurice, S. (2007). Historical pragmatics: what it is and how to do it. In S. Fitzmaurice & I. Taavitsainen (Eds.), *Methods in historical pragmatics* (pp. 11-36). Berlin: De Gruyter Mouton.

Taavitsainen, I. & Pahta, P. (2010). *Early Modern English Medical Texts: corpus description and studies.* Amsterdam: John Benjamins.

Tagg, C., Baron, A. and Rayson, P. (2012). “I didn’t spel that wrong did i. Oops”: analysis and normalisation of SMS spelling variation. In L. Cougnon and C. Fairon (Eds.) *SMS Communication: A linguistic approach.* Special issue of Linguisticae Investigationes 35:2, pp.367-388.

Thomas, V. (2012). *Piers Plowman, legal authority, and the law of subject status.* Unpublished PhD thesis: Washington University in St. Louis.

Thomas, A. (2018). *Piers Plowman and the reinvention of church law in the Late Middle Ages.* Toronto: University Press.

Thompson, P. & Alba-Juez, L. (2014). *Evaluation in context*. Amsterdam: John Benjamins.

Threlfall-Holmes, M. (2005). *Monks and markets: Durham Cathedral Priory 1460-1520*. Oxford: University Press.

Tiersma, P. (1999). *Legal language.* Chicago: University Press.

Tognini-Bonelli, E. (2001). *Corpus linguistics at work.* Amsterdam: John Benjamins.

Turner, M (1987). *Death Is the Mother of Beauty: Mind, Metaphor, Criticism*. Chicago: University of Chicago Press.

Turville-Petre, T. (1977). *The alliterative revival*. Cambridge: D.S. Brewer.

Turville-Petre, T. (2013). Alliterative horses. *Journal of English and Germanic Philology*, 112(2), 154-178).

Vallins, G.H. & Scragg, D.G. (1965). *Spelling*. London: Andre Deutsch.

Van Gelderen, E. (2018). *Analysing syntax through texts: Old, Middle, and Early Modern English.* Edinburgh: University Press.

Verdonk, P. (2002). *Stylistics.* Oxford: University Press.

Verdonk, P. (2013). *The stylistics of poetry: context, cognition, discourse, and history.* London: Bloomsbury.

Wales, K. (2001). *A dictionary of stylistics.* Second edition. Harlow: Pearson Education Ltd.

Wales, K. (2014a). *A dictionary of stylistics*. Third edition. Harlow: Pearson Education Ltd.

Wales, K. (2014b). The stylistic tool-kit: methods and sub-disciplines. In P. Stockwell & S. Whiteley (Eds.), *The Cambridge handbook of stylistics* (pp. 32-45). Cambridge: University Press.

Walker, B. (2010). Wmatrix, key concepts, and the narrators in Julian Barnes' Talking it Over. In B. Busse & D. McIntyre (Eds.) *Language and style*. Basingstoke: Palgrave, pp. 364-387.

Wallage, P. (2017). *Negation in Early English: grammatical and functional change.* Cambridge: University Press.

Weiskott, E. (2016). Piers Plowman and the durable alliterative tradition. *Yearbook of Langland Studies* 30, pp.123-173.

Weiskott, E. (2017). *English alliterative verse: poetic tradition and literary history.* Cambridge: University Press.

Werth, P. (1977) 'The linguistics of double-vision', *Journal of Literary Semantics* 6, pp. 3-28.

Werth, P. (1994). Extended metaphor – a text-world account. *Language and Literature* 3(2), 79-103.

Werth P. (1999) *Text Worlds: Representing Conceptual Space in Literature*. Harlow: Longman.

Whitelaw, C., Garg, N., & Argamon, S. (2005). Using appraisal taxonomies for sentiment analysis. *Proceedings of the First Computational Systemic Functional Grammar Conference*. University of Sydney, Australia.

Wilkins, K. (2018). *Characterization in Ælfric's Esther: a cognitive stylistic investigation.* Unpublished PhD dissertation, University of Nottingham.

Williams, R. (1983). *Keywords: a vocabulary of culture and society,* 2nd edition, London: Fontana Press.

Wittig, J. (2001). *Piers Plowman concordance*. New York: The Athlone Press.

Wittig, S. (1978). *Stylistic and narrative structures in the Middle English romances*. Austin: University of Texas Press.

Woolgar, C. (1995). Diet and consumption in gentry and noble households: a case study from around the Wash. In R.E. Archer & S. Walker (Eds.), *Rulers and ruled in late medieval England: essays presented to Gerald Harris.* London: Hambledon Press, 17-31.

Woolgar, C. (2001). Fast and feast: conspicuous consumption and the diet of the nobility in the fifteenth century. In M. Hicks (Ed.), *Revolution and consumption in late medieval England.* Woodbridge: The Boydell Press, 7-26.

Woolgar, C., Serjeantson, D., & Waldron, T. (2005). *Food in medieval England: diet and nutrition*. Oxford: University Press.

Xiao, R. (2015). Collocations. In D. Biber and R. Reppen (Eds.), *The Cambridge handbook of corpus linguistics*. Cambridge: University Press, pp. 106-124.

Yamamato, M. (1999). *Animacy and Reference: A Cognitive Approach to Corpus Linguistics.* Amsterdam: John Benjamins.

Yamamoto, M. (2006). *Agency and Indefiniteness: Their Linguistic and Cultural Manifestations.* Amsterdam: John Benjamins.

Yazdanjoo M, Sabbagh M and Shahriari H (2016) Stylistic features of Holden Caulfield's language in J.D. Salinger's The Catcher in the Rye: a corpus-based study. *English Studies* 97(7): 763-78.

Zunshine, L. (2003). Theory of Mind and experimental representations of fictional consciousness. *Narrative* 11, pp.270-91.

Zunshine, L. (2006). *Why we read fiction: Theory of Mind and the novel*. Columbus: Ohio State University Press.

APPENDIX A: TABLES, KEY WORD LISTS, AND CONCORDANCES

Table 3.2: List of the *ARCorp* source-texts and their databases of origin

Title	**Digitized edition used for *ARCorp***	**Source of the digitized edition**
The alliterative Morte Arthure (The Death of Arthur)	Benson & Foster (1994)	TEAMS
The Awntyrs of Arthure	Hahn (1995)	TEAMS
Cleanness (Purity)	Morris (1869a)	Project Gutenberg
The Destruction of Troy	Panton & Donaldson (2006)	Corpus of ME Prose and Verse
Gologras and Gawain	Hahn (1995)	TEAMS
Joseph of Arimathie (The Romance of the Holy Grail)	Skeat (1871)	Corpus of ME Prose and Verse
Mum and the Sothsegger	Dean (2000)	TEAMS
The Parlement of the Thre Ages	Ginsberg (1992)	TEAMS
Patience	Morris (1869a)	Project Gutenberg
Pearl	Morris (1869a)	Project Gutenberg
Pierce the Ploughman's Crede	Dean (1991)	TEAMS
The Pistel of Swete Susan	Peck (1991)	TEAMS
Richard the Redeless	Dean (2000)	TEAMS
Saint Erkenwald	Gollancz (1926)	Oxford Text Archive
The Siege of Jerusalem	Livingston (2004)	TEAMS
Sir Gawain and the Green Knight	Tolkien & Gordon (1967)	Corpus of Middle English Prose and Verse
The Wars of Alexander (Alexander Fragment C)	Skeat (1886)	Corpus of ME Prose and Verse

Wynnere and Wastoure	Ginsberg (1992)	TEAMS

Table 3.3: List of the *MVCorp* source-texts and their databases of origin

Title	**Digitized edition used for *ARCorp***	**Source of the digitized edition**
Anelida and Arcite	Skeat (1923a)	Project Gutenberg
The Book of the Duchess	Skeat (1923a)	Project Gutenberg
The Canterbury Tales	Skeat (1923d)	Project Gutenberg
Confessio Amantis	Macauley (1900)	Project Gutenberg
The House of Fame	Skeat (1923c)	Project Gutenberg
The Legend of Good Women	Skeat (1923c)	Project Gutenberg
The Parlement of Foules	Skeat (1923a)	Project Gutenberg
The Plowman's Tale (The Complaynte of the Ploughman)	Bowers (1992)	TEAMS
The Romance of the Rose	Skeat (1923a)	Project Gutenberg
Troilus and Criseyde	Skeat (1923b)	Project Gutenberg

(on CD – the 855 KWs generated for *PPCorp,* as set out in Excel)