

A survey of grammatical variability in Early Modern English drama

Andrew Hardie
Lancaster University, UK

Isolde van Dorst
Vienna University of Economics and Business, Austria

Corresponding author: Andrew Hardie, Department of Linguistics and English Language
Lancaster University, Lancaster, LA1 4YL, UK
Email: a.hardie@lancaster.ac.uk

Abstract

Grammar is one of the levels within the language system at which authorial choices of one mode of expression over others must be examined to characterise in full the style of the author. Such choices must however be assessed in the context of an understanding of the extent of variability that exists generally in the language. This study investigates a set of grammatical features to understand their variability in Early Modern English drama, and the extent to which Shakespeare's grammatical style is distinct from or similar to that of his contemporaries in so far as these features are concerned. A review of prior works on Shakespeare's grammar establishes that the quantitatively informed corpus linguistic approach utilised in this study is innovative to this topic.

Using two of the grammatically annotated corpora created by the *Encyclopedia of Shakespeare's Language* project, one made up of Shakespeare's plays, one of plays by other playwrights of the period, we present a method which steers a course between the narrow focus of close reading and the naïvely quantitative metrics of authorship analysis. For a set of 15 grammatical features of stylistic interest, we retrieve all instances of each feature in each play via complex corpus search patterns and calculate its relative frequency. These results are then considered, in aggregate and at the text level, to assess the differences across plays, across dramatic genre, and between Shakespeare and the other dramatists, via both statistical summary and visual representation of variability.

We find that Shakespeare's grammatical style tends (especially in comedies and tragedies) to disprefer informationally-dense noun phrases relative to the other playwrights; and, moreover, to prefer tense, aspect and pronoun features which suggest a greater degree of narrative focus in his style. Furthermore, we find Shakespeare to be *highly* distinct in his preferences regarding verb complement subordinate clause types. These findings point the way both to a novel methodology *and* to further as yet unconsidered questions on the subject of Shakespeare's grammatical style.

Keywords

grammar, Shakespeare, variability, Early Modern English, corpus, style

1. Introduction

The aim of this study is to investigate the grammatical style of Shakespeare and of other dramatists of his period,¹ in terms not of qualitative distinctions but of quantitative preferences regarding their use of a range of specifically grammatical features. We focus on variability,² aiming to establish the

extent to which the incidence of different features varies across different plays, and the extent to which Shakespeare is distinct from or similar to his contemporaries in terms of this variation in grammatical style.

The role of grammar in style is appreciated more in some quarters than others (specifically, within different subdisciplines of linguistics, as opposed to literary studies). In corpus linguistics, it is a commonplace of keyword analysis that lexical keywords tend to reflect content whereas grammatical keywords reflect style. In a keyword comparison of, say, spoken and written English, the typical prominence of keywords such as *I* and *you* (spoken) and *the* and *of* (written) is interpreted as reflecting the style-related differences between these two modes of language. In other words, grammatical elements are held to be at the core of style. Likewise, stylometric authorship attribution methods are typically based on word frequencies, looking at the most frequent (i.e. grammatical) words in order to distinguish the stylistic 'fingerprints' of different authors.

Within stylistics specifically, grammar as a marker of style has received considerable attention throughout the long history of the field (see, for instance, Crystal and Davy, 1969; Enkvist, 1973; Short, 1996; Jeffries and McIntyre, 2010; Stockwell, 2020). To consider a single example of this attention, we may note that the highly influential study of Leech and Short (1981) addresses grammar as a focus of stylistic analysis with detail and systematicity. They note that:

If we define style in terms of stylistic variants, we assume that language specifies a repertoire or code [of] possibilities, and that a writer's style consists in preferences exercised within the limits of the code.

(Leech and Short, 1981: 138)

Grammar is then one of the levels within the linguistic code at which choices among possibilities must be made, choices that constitute the style of the work or the author. Leech and Short's exposition of a 'checklist' of features which may be relevant in a stylistic analysis (Leech and Short, 1981: 75-80) includes a large set of grammatical categories, features, and constructions, from across the whole system of English grammar: word classes, phrase structure, clause structure, and sentence structure. For instance, they suggest considering, with regard to just clause types, the following:

What types of dependent clauses are favoured: relative clauses, adverbial clauses, different types of nominal clauses (*that*-clauses, *wh*-clauses, etc)? Are reduced or non-finite clauses commonly used, and if so, of what type are they (infinitive clauses, *-ing* clauses, *-ed* clauses, verbless clauses)?

(Leech and Short, 1981: 77)

The words *favoured* and *commonly* imply that the assessment should be quantitative, a point that the authors indeed make explicitly elsewhere (Leech and Short, 1981: 43ff, 111-117, *passim*). By making a list of relevant features and then suggesting that counting their frequencies (and identifying features whose relative frequency in a text under analysis is exceptional) lies at the base of analysis, Leech and Short anticipate the equally influential work of Biber (1986, 1988). Biber's *multi-dimensional analysis* method employs rather more complex statistical techniques, but begins with the enumeration of a list of features, and then the measurement of their frequency in a text or texts. Biber's method largely automates the process of counting by using specialised corpus searches to identify all examples of each feature, but on a conceptual level the process is very similar. Consider, for instance, the following features from Biber's list, which cover the area of dependent clauses on which we quoted Leech and Short above:

21. *that* verb complements (e.g., *I said that he went.*)
 22. *that* adjective complements (e.g., *I'm glad that you like it.*)
 23. WH clauses (e.g., *I believed what he told me.*)
 24. infinitives
 25. present participial clauses (e.g. *Stuffing his mouth with cookies, Joe ran out the door.*)
 26. past participial clauses (e.g., *Built in a single week, the house would stand for 50 years.*)
 - [...]
 29. *that* relative clauses on subject position (e.g., *the dog that bit me*) [...]
 31. WH relatives on subject position (e.g. *the man who likes popcorn*) [...]
 35. causative adverbial subordinators (*because*) [...]
 38. other adverbial subordinators (e.g. *since, while, whereas*)
- (Biber, 1988: 73-74)

The similarity should require no further comment.

This study, then, draws on both the view of grammatical style outlined by Leech and Short, and the corpus-based quantitative methods for identifying and quantifying linguistic features introduced by Biber. It is structured as follows. In section 2, we survey existing work on the grammar of Shakespeare's drama, demonstrating that an analysis from the perspective we adopt fills an identifiable gap in the field. Section 3 outlines our methodology, first in overview, and then with precise detail of the corpora and analytic procedures utilised. The data generated by these procedures is presented, and discussed, in section 4; finally we draw together our findings in a brief conclusion (section 5).

2. Previous work on the grammar of Shakespeare's drama

The analyses which we present in the second half of this paper draw on established corpus linguistic and stylistic approaches, but are, we propose, an innovative contribution in the context of the study of Shakespeare's language. That is to say, there exists a gap in the literature to date for such a quantitatively oriented analysis. To support that claim, in this section we review the literature in question and demonstrate that, despite the variety of prior work in this area, there does indeed exist a lacuna for a quantitative, corpus-based analysis of grammatical style in the plays of Shakespeare.

Prior work on the grammatical style of Shakespeare's dramas³ can be classified under three broad headings. Under the first are book-length works which seek to provide a comprehensive account of all features relevant to the issue of grammatical style; two of these, Blake (2002a) and Hope (2003) will be reviewed in some detail below. The second heading covers works on Shakespearean language or style which address Shakespeare's grammar in passing but are not primarily studies of grammar. Under the third heading are studies, often published in stylistics journals, which explore one or a small number of specific features of the grammar of Shakespeare without attempting to be exhaustive.

Starting with the third category, one area which has attracted significant attention is the use in Shakespeare of second person pronouns. The *thou/you* distinction of the time is a matter of perennial debate: as the distinction would become archaic within about a century, and at the time was driven as much by social status and politeness as by grammatical number, it is both of historical interest and important as a pragmatic device exploitable for dramatic purposes by a writer such as Shakespeare. Notable among the studies of *thou/you* are U. Busse's (2002) monograph, as well as papers by, among others, Mulholland (1967), Wales (1983), Mazzon (2003), Stein (2003) and van Dorst (2018). In an area with similar pragmatic relevance, B. Busse (2006) examines vocatives in Shakespeare, defined as "direct attitudinal adjunct-like forms of address[,] [r]ealised as a nominal group or head" (B. Busse,

2006: 26-27). Terms of address, similarly defined, have been studied by U. Busse (1998), Blake (2002b) and Jucker (2012). Other grammatical features which have attracted attention include part-of-speech category conversion (Wales, 1987); complement clauses (Fanego, 1990a, 1990b); auxiliary and copula verbs (Kakietek, 1987; Biese, 1987); third person present tense verb forms (Taylor, 1987); and the colloquial use of *your* as a generic deictic determiner (Wales, 1985).

Only somewhat at home in this third category is Shore (2015). Shore examines three exemplars of grammatical *constructions* (in the sense of the Construction Grammar theory of Goldberg, 1995). The three are⁴ (a) *what (article) (up to 5 tokens) (form of 'be') I*, as in Hamlet's 'what a rogue and peasant slave am I!'; (b) *(past participle), (past participle), (optional 'and'/or) (past participle)*, as in Hamlet's father's 'Unhousel'd, disappointed, unaneled'; and (c) *to X or not to X*, as in, of course, 'To be or not to be' (Shore, 2015: 117,122,130). Using corpus techniques to identify additional instantiations of these constructions in Shakespeare and in the *Early English Books Online* collection puts Shore within the tradition of research on a few narrowly focused facets of Shakespeare's grammar; yet Shore's purpose is, in fact, to argue for the necessity of a *constructicon* (i.e. lexicon of constructions) in which such constructions could be recorded and analysed instead of a grammar of descriptive or generative rules.

Let us now move to the second category, of works on Shakespeare's language which are largely non-grammatical in focus. Books on the language of Shakespeare for both scholarly and popular audiences are numerous (the latter type, which we do not consider further, is exemplified by Kermode, 2000), and we do not claim that this review is at all comprehensive. We merely consider some examples in which the minimal attention paid to grammar relative to other aspects of language or style is notable. One straightforward example is Alexander's (2004) retrospective collection of sixteen language-related papers from the yearbook *Shakespeare Survey*, which includes *no* contributions that deal primarily with grammatical questions. Similarly, of the eleven contributions to Ravassat and Culpeper's (2011) edited collection on Shakespearean stylistics, only one is oriented primarily to grammar – Ingham and Ingham (2011) on Shakespeare's use of subject-verb inversion relative to the rhythm of iambic pentameter. Grammar is less underrepresented in Salmon and Burness (1987), among whose 33 chapters are 11 that relate, in one way or another, to grammar (some examples of which were mentioned earlier in this review).

Moving from edited collections of research papers to monographs and textbooks, we see a tendency for grammar to be addressed in one chapter among many. Johnson's (2013) textbook on, and his (2019) historical survey of research into, the language of Shakespeare, both proceed in this manner, as do volumes by Blake (1983) and Hope (2010).

Blake (1983: vii) explicitly disclaims the goal of presenting a full grammar of Shakespeare's language, rather declaring his aim to be "to help readers come to a greater understanding of the language of the time" while trying "to avoid using linguistic jargon". Yet in fact rather more than half of the book (pp. 56-130) is devoted to grammar; and, in our view, his treatment of Shakespeare's grammar does in fact approach being exhaustive. Blake's avoidance of "jargon" may sometimes render these accounts suboptimal for the linguistically-trained reader, but it certainly enhances accessibility for the lay reader.

Hope's (2010) volume on Shakespeare's language is centrally concerned with drawing out the implications of the differences between Renaissance and modern views on language. He devotes a chapter to syntax (Hope 2010: 138-169) which focuses on the stylistic or rhetorical effects of certain grammatical phenomena – the manipulation of subject/agent status, for instance. These devices are according to Hope particularly common in Shakespeare's late style.

Johnson (2013) frames grammar in Shakespeare as a “half-way house” between Old English as a synthetic language and contemporary English as an analytic language (evidently a simplification for pedagogic reasons). So, for instance, his discussion of the noun phrase includes split genitives (Johnson, 2013: 89) and irregular plurals (p. 90); his discussion of the verb phrase includes the process by which third singular *-(e)th* was ultimately replaced by *-(e)s* (Johnson, 2013: 95-101). Each point addressed is used additionally as an opportunity to present relevant introductory material in linguistics (e.g. Johnson, 2013: 82-84 on inflections; pp. 96-98 on how language changes spread through society). More technical than Hope (2010) but less exhaustive than Blake (1983), this chapter is in our view the best available short introduction to the issues in question.

In addition to their monographs on Shakespeare's language, both Hope and Blake have published volumes on Shakespeare's grammar specifically. These are the two works which we characterised (above) as attempts to deal comprehensively with the subject – the first such for many decades, according to both authors. The most notable grammar to predate that hiatus is that of Abbott (1870), whose legacy looms large over both Hope's and Blake's grammars in different ways. Hope (2003: 1-2) reports that his grammar was commissioned as a replacement for Abbott's, and at several points, refuting Abbott's (typically Victorian) prescriptive stance emerges as Hope's central priority. One instance among many is a discussion of “anomalous pronouns” in which Hope (2003: 93ff) observes that:

Abbott's prescriptivist approach, and his interest in illustrating the differences between Early Modern and nineteenth-century English, mean that he spends a long time listing instances of ‘anomalous’ pronouns. His stance on such forms is made clear by terms like ‘misused’ and ‘ungrammatically’ (205). Modern linguists would tend to see variation in pronoun form as part of the natural variation of the language [...]

Hope takes pains in these cases to emphasise that what Abbott deemed objectionable is merely expected variation. In discussing such variation Hope often incorporates information on frequency, that is, how often one variant as opposed to the other(s) is used in the Shakespeare corpus. When he contextualises this information by discussing what was common or usual in Early Modern English outside Shakespeare, however, similar quantification is not provided (most likely because of the lack at the time of suitable data on which to base any such quantification).

Like Hope, Blake (2002a) notes the persistence of Abbott (1870) as the primary reference for Shakespeare's grammar. However, for Blake the consequent problem is the lack of attention paid to linguistic issues by scholars responsible for modern editions of Shakespeare. Early on, Blake (2002a: x) observes that “[t]he language of Shakespeare is a topic to which literary scholars pay lip-service, though they rarely pay as much attention to it as it deserves”. He later explains this state of affairs as follows:

[...] modern editors and critics of Shakespeare's works [...] often find modern linguistics difficult or irrelevant to their concerns, and so they have in their turn become less interested in the minutiae of the language of the text they are editing. Often a mere reference to Abbott or to the *Oxford English Dictionary* satisfies their conscience that they have explained the difficulty in syntax or vocabulary that exists in the text. [...] Editors spend more of their time grappling with other problems, such as the printing of [the First Folio] or individual quartos, or the cultural references in or the productions of the play they are dealing with, and thus fail to offer satisfactory explanations of the language which may baffle many readers.
(Blake, 2002a: 11)

This theme emerges throughout, as Blake comments on decisions made in modern editions which seemingly reflect ignorance of some aspect of Early Modern English grammar. For instance, in a section on subordinate conjunctions, Blake (2002a: 168-169) notes one edition's punctuation of a line in *Macbeth* so as to rule out a reading of conjunction *that* in a function it no longer possesses today, a reading that Blake considers preferable.

The fundamental organisation of Blake (2002a) is clearly based on the structure of the grammar-related chapters of his earlier (1983) volume. Blake achieves comprehensiveness by moving through each facet of the grammatical system point-by-point, exploring every topic in depth before moving to the next. So, for instance, a section on prepositions (pp. 177-200) begins with a number of preparatory matters, before a subsection discussing each preposition in alphabetical order; for each, a list of meanings or functions is enumerated, with multiple examples from the plays for every function. It would not be inappropriate to liken Blake's approach and level of detail to that of descriptive grammars written by field linguists after the exhaustive point-by-point questionnaire of Comrie and Smith (1977). Largely missing, however, is quantification. For example, one entry in the aforementioned list of prepositions is "*Toward*, occasionally *towards*" (Blake, 2002a: 195), with no figures placed on the rarity of the latter. In fact, the variation between *toward/s* is fairly evenly split, with 102 examples of the former and 76 of the latter in the plays of Shakespeare.⁵ *Towards* is much more than occasional. Despite overlooking the role of quantification here, Blake (2002a: xii) does note the potential of corpus-based analyses for contextualisation of Shakespeare's language relative to that of his time.

Hope (2003: 9) is clearly correct to observe that he and Blake had "very different approaches, with different audiences in mind". However, as we have seen, they are alike in that they intermittently exhibit awareness of the importance of quantification to aspects of their topic, whilst however not integrating quantitative analysis into their accounts, likely due to unavailability of the data and methods that would have allowed them to do so. Thus, even across the combination of both these monographs, there exists as yet no detailed quantitative or corpus-based analysis of grammatical features of style in Shakespeare in the context of the English of his time that is even remotely akin to what Biber et al. (1999), for example, have achieved for present-day English. We hope our analyses in this paper, directed *specifically* to quantifying variation in use of grammatical features, will contribute to filling this gap.

3. Method

3.1. Overview: two points of contrast

As previously indicated, our perspective on the matter of grammatical style is rooted in corpus-based analytic methodologies. In many cases, corpus-based studies of literary style have availed themselves of complex analytic frameworks from the study of style, discourse or rhetoric, whilst not using more than fairly simple corpus methodologies. For instance, in a study of James Joyce's short story 'Eveline', following Stubbs (2001), O'Halloran (2007) uses the keywords method and then compares the frequencies of particular patterns whose salience is indicated from that keywords analysis. In studies such as O'Halloran's, the overall number of instances of any particular item of interest is typically small; e.g. following the appearance of *would* as a keyword in the text he analyses, O'Halloran looks at 18 examples of *would* within an expression of free indirect thought. The virtue of covering only a small number of examples is that it permits the utilisation of analytic techniques which require a very close reading of the full dataset, and which are thus in practice not applicable to larger scale data. For instance, O'Halloran deploys a Hallidayan transitivity analysis across these 18

instances of *would* that would have been quite impossible to apply to (say) 2,000 examples, but which is revelatory to O'Halloran's investigation.

Our purpose here is quite different. We aim to draw on a rather wider range of corpus methods, most especially, the advanced corpus queries permitted by the use of part-of-speech annotation and lemmatisation that reference features more abstract than the actual words of the text, in order to note the location and frequency of instances of a large set of grammatical phenomena. That step complete, our procedure is then *not* to look in depth at the discourse-pragmatic function in context of this or that particular device but rather to allow differences of frequency across plays within the First Folio, and likewise across plays by Early Modern writers other than Shakespeare. The aim of this procedure is to reveal quantitative differences – that is, differences of (conscious or subconscious) authorial *preference* – in the use of the grammatical repertoire of English overall. Our goal could be informally characterised as a bird's-eye overview of the waves of the grammatical-stylistic ocean, rather than a sea-level snapshot or submarine deep dive.

But that stance, and our use of statistical techniques, requires us on the other hand to distinguish our approach from that of stylometry. The apparatus of stylometry is quantitative, but the features used to measure stylistic difference, that is frequencies of the most frequent 50 or 100 words, are pre-theoretical (or perhaps atheoretical), as are the units of analysis – typically chunks of 500 sequential words ignoring scene, act and play boundaries. In the context of Shakespearean studies, stylometry is often utilised for authorship attribution, attempts to ascertain the true writer of (parts of) plays of disputed Shakespearean authorship. A good recent illustration of applying these methods to the Shakespeare plays is that of Craig (2018), who (like us) uses a corpus of plays by other writers alongside and in comparison to Shakespeare's works. Craig, to his credit, avoids a common pitfall in stylometry where *only* the statistical constructs generated from the data, and not the textual behaviour of the underlying high-frequency words, are seen to be of interest. On the one hand, then, Craig's initial analysis uses the technique of Linear Discrimination Analysis to generate a function which scores text chunks based on word-frequency weightings extracted algorithmically so as to maximally separate the Shakespeare and non-Shakespeare chunks (Craig, 2018: 170-172). But on the other hand he *also* undertakes a qualitative discussion of words identified as highly characteristic or counter-characteristic of Shakespeare's style. For instance, Craig (2018: 173) identifies *the*, *which*, *hath*, *thou*, *you* and *that* as words characteristic of Shakespeare relative to others, and conversely *all*, *now*, *can*, and *any* as characteristically dispreferred by Shakespeare. Of the latter elements, Craig notes that they are "used most often in the plays for emphasis and as filler words"; and after a more detailed qualitative analysis of *the* and *all* in context concludes:

Shakespeare's dialogue has relatively frequent recourse to 'the', and uses 'all' less than the run of his contemporaries. The first suggests his writing is more concrete, and the second that he is less prone to filler words and to sweeping generalisations.
(Craig, 2018: 177)

Although the naïve processes used in stylometric methodology are entirely by design, we do not consider them appropriate for our purposes. In contrast, then, our analysis respects the plays as the units across which stylistic difference shall be assessed, and utilises a method that does not ignore but rather builds on existing knowledge of abstract grammatical patterns and categories.

3.2. Data

The two corpora used in this study are the *Enhanced Shakespearean Corpus: First Folio Plus* (hereafter, *ESC:Folio*) and the *Enhanced Shakespearean Corpus: Comparative Plays* (hereafter,

ESC:Comp). Both resources are openly available (at present via Lancaster University's CQPweb server and in future as full downloads)⁶ under a Creative Commons licence. *ESC:Folio* contains 38 plays – the 36 in the First Folio, with the addition of *The Two Noble Kinsmen* and *Pericles Prince of Tyre*. *ESC:Comp* contains 46 plays by 24 other playwrights from the same time period as Shakespeare. Both corpora are just over 1 million words in length, and are comparable in terms of their genre distribution among the plays. The composition of the two corpora is discussed in full in Culpeper et al. (forthcoming) for *ESC:Folio* and Demmen (2020) for *ESC:Comp*.

Among the many mark-up and annotation enhancements applied to *ESC*, we rely on (i) text boundaries and textual metadata for each play – notably the plays' genre (comedy history, or tragedy); (ii) part-of-speech annotation; (iii) lemmatisation; and (iv) spelling regularisation, which improves search retrieval rates as well as being prerequisite to (ii) and (iii). The part-of-speech annotation in *ESC:Comp* was applied automatically with the CLAWS tagger,⁷ which is usually around 95-97% accurate, but likely to perform less well on Early Modern drama. The CLAWS-generated annotation in *ESC:Folio* has been manually post-edited throughout, so that it is as close to 100% correct at the major part-of-speech level as is humanly possible (a novel accomplishment for digitally annotated versions of Shakespeare; see Culpeper et al., forthcoming).

3.3. Procedures

Our procedures were designed to generate a quantitative, corpus-based analysis characterising high-level features of grammatical style and any differences therein between Shakespeare and his contemporaries, while also considering genre contrasts (among comedies, tragedies, and histories). The features under analysis form part of both Biber's (1988) list of linguistic features and Leech and Short's (1981) list of grammatical style features, and are also addressed within the Shakespeare grammars of Blake (2002a) and Hope (2003) *inter alia*. These features were then considered from the perspective of diversity of grammatical style in Shakespeare and the other represented playwrights.

Thus, critically, we did not limit our procedure to an overall "Shakespeare *versus* everybody else" comparison. While not overlooking differences between the two corpora as a whole, our primary focus is on the variability of particular grammatical feature frequencies across plays as well as the distribution of features across genres of play.

Grammatical variation at the grossest level may be captured in the relative frequencies of major part-of-speech categories (see Rayson et al., 2002; cf. Hardie, 2007). Therefore, the first six features under analysis were verbs, nouns, adjectives, adverbs, pronouns, and determiners as undifferentiated categories. But since the early 1980s research in both stylistics and corpus linguistics has identified numerous more narrowly defined grammatical phenomena the rate of whose presence in a text is (at least potentially) revelatory of style (see section 1 above). It is beyond the scope of any single study to encompass *all* such phenomena; we selected nine as additional features of interest (listed in table 2). Rather than argue for the relevance of these particular phenomena here, where necessary we supply some background comments in the course of discussing each feature in section 4 below.

The CQPweb software (Hardie, 2012), within which all analyses were undertaken, is a browser-based interface to the Corpus Workbench suite of programs and in particular CQP ('Corpus Query Processor'), a query engine that supports powerful phrase-level regular-expression and annotation-based corpus searches. Of the two query languages available in CQPweb, we used the full CQP syntax (see Evert and Hardie, 2011). The relevant features of this query language are noted below.⁸

For each feature under analysis, we found or devised a CQP query that would retrieve all, or nearly all, instances of that feature from the corpus. The nine features beyond major part-of-speech are all among the larger feature-set utilised in Biber's (1988) landmark work, and the query definitions were very much driven by the descriptions given in the methodological appendix of Biber (1988), as table 2 details. The development of these CQP queries for the Biber features is related by Hardie et al. (forthcoming); the only alterations we made here were to allow for the Early Modern English version of the part-of-speech tagset, which adds extra tags for second person singular forms – such as VMT for *wouldst*. That accomplished, we (a) ran the query on each of the corpora; (b) retrieved the absolute and relative frequencies of the feature for each of the genres and individual plays across both corpora; and (c) generated plots of the frequencies per play allowing (qualitative) interpretation and assessment of the (quantitative) diversity – the *ranges* of frequencies across plays, across genres, and across the two corpora representing Shakespeare and the other playwrights. The end-product plots are presented and discussed in the following section; tables 1 (for major part-of-speech) and 2 (for other phenomena) list the features, their definitions, and their expression as CQP queries, together with the absolute and relative frequencies in the two corpora considered as a whole. We express relative frequencies as counts per million words (PMW). PMW is often suboptimal as the basis for calculating relative frequency for things other than words, but it has the critical advantage that it can be calculated for *all* the features we are looking at, which is not true for other possible bases (e.g. frequencies per *N* verb phrases, per *N* sentences, per *N* subordinate clauses).

Feature	Basis of query	CQP query	Abs. freq.		Rel. freq. (PMW)	
			<i>Folio</i>	<i>Comp</i>	<i>Folio</i>	<i>Comp</i>
Verbs	Words tagged as verbs.	[pos="V.*"]	187,519	200,135	180,533	183,387
Nouns	Words tagged as nouns.	[pos="N.*"]	176,991	198,843	170,397	182,203
Adjectives	Words tagged as adjectives.	[pos="J.*"]	49,923	51,821	48,063	47,484
Adverbs	Words tagged as adverbs.	[pos="R.*"]	58,309	63,367	56,137	58,064
Pronouns	Words tagged as pronouns.	[pos="P.*"]	100,228	100,726	96,494	92,297
Determiners	Words tagged as determiners.	[pos="D.*"]	32,178	35,664	30,979	32,679

Table 1. Queries to extract instances of major parts-of-speech.

Feature	Basis of query	CQP query	Abs. freq.		Rel. freq. (PMW)	
			<i>Folio</i>	<i>Comp</i>	<i>Folio</i>	<i>Comp</i>
Past tense verbs	Words tagged as such.	[pos="V.D.?"]	16,034	15,958	15,437	14,623
Perfect aspect verbs	Verb <i>have</i> , followed by past participle (with optional intervening adverb(s) and interrogative subject).	[pos="VH.*"] [pos="(N A?PP).*"]? [pos="[RX].*"]* [pos="V.N.?"]	6,064	5,559	5,838	5,094
Present tense verbs	Words tagged as base form (includes imperative/present subjunctive) or as third-singular or second-singular form	[pos="VBM VBR[V[BDHV][0ZT]"]	83,490	90,589	80,380	83,008
1st person pronouns	Words tagged as such (PPI*), or matching a reflexive form, or <i>mine</i> as pronoun (not noun/verb).	[pos="PPI.." word="my(self)? our(s self selves)?"]%c taglemma="mine_PRON"%c]	53,553	56,160	51,558	51,460
2nd person pronouns	Words tagged as such (PPY*), or matching a reflexive form.	[pos="PPY.*" word="your(s self selves)? th(ou e y yself ine(self)?"]%c]	35,402	37,240	34,083	34,124

3rd person pronouns	Words tagged as such (PPH*), excluding <i>it</i> , or matching a reflexive form.	[pos="PPH." word="her(s self)? hi(s m(self)?) the(m irs?)(self selves)?"%c]	33,557	33,701	32,307	30,881
<i>That</i> -clause verb complements	Words tagged CST for complementiser <i>that</i> , plus surrounding possible patterns as defined by Biber (1988:230) using relevant semantic verb classes.	([word="and nor but or also"%c pos="Y.*"] [pos="CST"] [(pos="A[^P].*" & word!="no"%c) (word="th(at is [oe]se)"%c & pos="D.*"%c) word="each all every many much few several some any"%c pos="MC(1 MC)?" pos="(A?PP.*EX N.2.*NP.* NNB)"] [taglemma="(acknowledge admit agree anticipate arrange ask assert assume beg believe claim command complain conclude decide declare demand demonstrate deny determine discover doubt estimate explain fear feel find forget grant guess hear hint hope imagine imply indicate infer insist instruct know learn mean mention notice ordain pledge proclaim promise pronounce propose protest prove reali[sz]e recogni[sz]e recommend remark remember reply report request reveal say see show stipulate suggest suppose swear think understand urge write seem appear)_V.*"%c] ((pos="I.*") [pos!="N.*"]* [pos="N.*"])? [pos="CST"])	2,263	1,789	2,179	1,639
Infinitives	Words tagged as TO (infinitive marker); limit of this feature to marked infinitives (as opposed to those, e.g., after modals) follows Biber (1988: 232)	[pos="TO"]	9,941	12,019	9,571	11,013
WH-clause verb complements	A <i>wh</i> -word followed by a clause and preceded by a verb of a relevant semantic class (following Biber 1988: 231).	[taglemma="(acknowledge admit agree anticipate arrange ask assert assume beg believe claim command complain conclude decide declare demand demonstrate deny determine discover doubt estimate explain fear feel find forget grant guess hear hint hope imagine imply indicate infer insist instruct know learn mean mention notice ordain pledge proclaim promise pronounce propose protest prove reali[sz]e recogni[sz]e recommend remark remember reply report request reveal say see show stipulate suggest suppose swear think understand urge write)_V.*"%c] "wh(om? ose ich) what where when how whether why whoever whomever whichever wherever whenever whatever however"%c [pos="[^V].* VV.*"]	469	569	452	521

Table 2. Queries to extract instances of complex grammatical features.

The main elements of CQP syntax used in tables 1 and 2 are as follows:

- [] : A set of conditions for matching a single token; these units are joined into phrase-level patterns with standard regular expression syntax (brackets for grouping, pipe for alternatives, and repetition markers).
- " " : A regular expression matching a single wordform or annotation; CQP uses Philip Hazel's *Perl Compatible Regular Expressions* library (PCRE).⁹
- %c : Flags a preceding regular expression as case-insensitive.
- **word=** , **pos=** , **taglemma=** : Labels that indicate which component of the token data a following regular expression is to be matched against; *word* is the wordform, *pos* is the CLAWS C6 part-of-speech tag¹⁰ (with Early Modern English extensions), *taglemma* is the

lemma annotation joined with an underscore to a simplified POS tag (major word classes only, as introduced in the BNC1994 XML Edition); if a quoted regular expression occurs outside square brackets, it is matched against *word*.

4. Analysis

4.1. From initial quantification of the corpora to a genre-aware analysis of texts

Let us begin from the highest-level standpoint possible: Shakespeare versus the other playwrights. For this purpose, we consider the relative frequencies from tables 1 and 2 for ESC:Folio and ESC:Comp as collections. Since the base frequencies of the phenomena in question vary so much – a few hundred PMW for WH-clauses, over 180,000 PMW for all verbs – direct comparison of these relative frequencies can only be done on a per-feature basis. To facilitate a full overview, we calculated a simple effect size statistic, the ratio of relative frequencies (RRF; the relative frequency in ESC:Folio divided by the relative frequency in ESC:Comp). These effect sizes are presented in table 3. An RRF of 1 indicates that a feature is about equally common in the two corpora; RRF below 1 indicates that the feature is rarer in ESC:Folio than in ESC:Comp; RRF above 1 indicates the opposite.

Feature	Effect	Interpretation
Verbs	0.98	
Nouns	0.94	Shakespeare's style <i>disprefers</i> nouns.
Adjectives	1.01	
Adverbs	0.97	
Pronouns	1.05	Shakespeare's style <i>prefers</i> pronouns.
Determiners	0.95	Shakespeare's style <i>disprefers</i> determiners.
Past tense verbs	1.06	Shakespeare's style <i>prefers</i> past tense.
Perfect aspect verbs	1.15	Shakespeare's style <i>prefers</i> perfect aspect.
Present tense verbs	0.97	
1st person pronouns	1.00	Frequency of 1st/2nd person pronouns appears not to discriminate Shakespeare's style from other playwrights.
2nd person pronouns	1.00	
3rd person pronouns	1.05	Shakespeare's style <i>prefers</i> 3rd person pronouns.
<i>That</i> -clause verb complements	1.33	Shakespeare's style <i>strongly prefers</i> this type of clause as verb object.
Infinitives	0.87	Shakespeare's style <i>disprefers</i> subordination via <i>to</i> -infinitive clauses.
WH-clause verb complements	0.87	Shakespeare's style <i>disprefers</i> this type of clause as verb object.

Table 3. Effect sizes (RRF to 2 d.p.) for differences in grammatical feature use between Shakespeare and other playwrights.

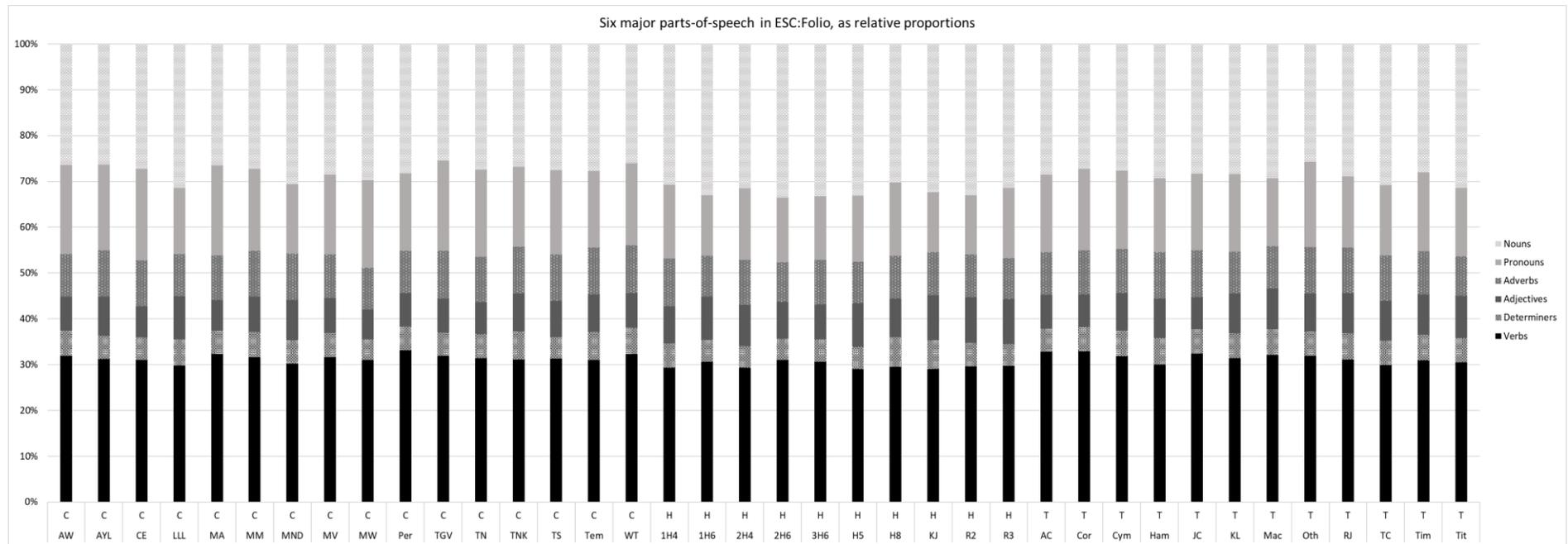
To focus attention on the major differences, table 3 applies the term *prefer/disprefer* to items with RRF of 1.05 or higher/0.95 or less respectively – as a shorthand for *makes relatively greater/lesser use of*. Even from this very small number of features, some of which are very coarse-grained, some consistencies which are linguistically interpretable can be observed. For instance, heavy noun phrases are an important locus of information in the clause. Given that, Shakespeare's dispreference for two of the building blocks of such phrases, that is nouns and determiners, and converse preference for pronouns, suggests a less informationally dense, and in consequence more speech-like, style (cf. Biber 1988; Rayson et al. 2002) than the average of the other playwrights. Going to the finer-grained features which discriminate pronouns by person, we see that the preference for pronouns in fact is only for third person pronouns. There seems to be no effect for first or second person pronouns – whose use is typically another feature of oral style. From the different complement subordinate clause features we see that, 'To be, or not to be' notwithstanding, *to*-infinitives are less used, and full clauses with *that* more used, in Shakespeare. Once again we note a difference in information density. Infinitive clauses are relatively more integrated into the main clause than are *that*-clauses, and thus, the main predicate is overall more informationally dense (it is not clear to us to what extent the same might apply to WH-clauses, a problem whose resolution is beyond the scope of this paper). The remaining points of interest are Shakespeare's preference for past tense and perfect aspect, which suggest – in light of the preference for third but not first/second person pronouns – that Shakespeare's style exhibits narrativity to a greater extent than the other playwrights. In sum, then, the overall comparison suggests that Shakespeare's grammatical style is differentiated from the average of the other playwrights by its narrativity and by its relative preference for lesser information density, but not for other features of orality.

Working with comparative frequency data such as that in tables 1 and 2, it may be tempting to apply statistical significance testing and to report a low p-value as establishing the relevance or importance of analyses such as the above. We have resisted this temptation. The two corpora yield at the word token level rather more than two million data points. Such vast absolute frequencies all but ensure that a null hypothesis significance test will produce an infinitesimal p-value that ultimately is not especially meaningful (see Cumming, 2014 for a technical discussion of this point). Being guided by effect size has allowed us instead to focus on the proportionally greatest differences in the foregoing initial exploration of Shakespeare's grammatical style.

That said, it is also worth highlighting what the examination of relative frequencies has *not* accomplished. It has *not* allowed us to account for variability in any way – because the overall relative frequencies are averages across all texts in each corpus. Neither have the plays' genres been taken into account, or the proportions of poetry and prose within different texts, or any other such internal heterogeneity. But it is exactly this variability which we aimed to explore to begin with. Again, one might be tempted to apply obvious statistical approaches – such as repeating the above analysis for comedies versus tragedies versus histories, or using multivariate regression statistics such as mixed-effect modelling (see e.g. Gries, 2015), or exploratory statistics such as factor analysis or cluster analysis. To our mind, however, much can be accomplished by finding ways to survey by eye the actual underlying data to understand what the variation across individual data points (that is, plays) is doing.

Figure 1 illustrates one way this can be done. In this chart, the relative proportions of the major parts-of-speech are shown vertically, with one stack per play, and the plays ordered left-to-right by genre first and then by the corpus ID code of the play.¹¹ Rendering the data thus, it is possible to get an immediate sense not only of the *level* of variability of these features across texts in the two corpora, but also the *locus* of the variation, and any differences in its patterning between the two corpora. To

begin with the obvious, there is rather *more* variation in ESC:Comp than ESC:Folio – almost certainly because the former includes many playwrights and the latter just one. The graph illustrates quite straightforwardly that the noun category (at the top of each stack) is an important locus of variation in both corpora; this reinforces the observations from the effect analysis. That pronouns seem to vary inversely to nouns, so that the combination of the two categories has more stable frequency than either alone, further underlines this. However, a point that could not be observed from the raw numbers is that the noun-pronoun balance appears to be conditioned by genre: the history plays in both corpora are generally more noun-heavy than the comedies and tragedies. We must then ask whether higher information density in histories might be behind the higher information density in playwrights other than Shakespeare. If ESC:Comp contained more historical drama than ESC:Folio, it would exhibit greater noun/determiner use for information density *even if* Shakespeare's stylistic preferences were no different from the other playwrights on this point. We must, therefore, consider the aspects of the corpus design¹² shown in table 4.



It may contain minor errors or infelicities; the publisher's typeset version (available at the foregoing link) is the only definitive form of the paper.

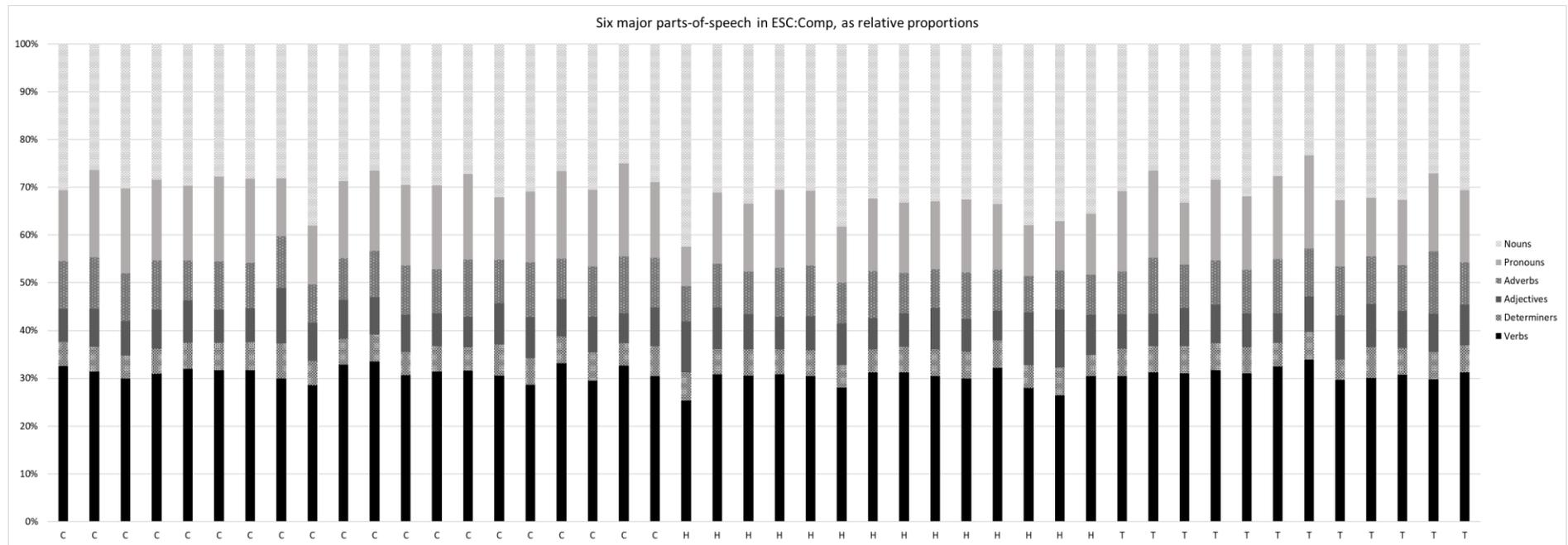


Figure 1. Major parts-of-speech shown proportionally by text, with plays grouped as comedies, histories and tragedies

Genre	ESC:Folio		ESC:Comp	
	Size in tokens	% of corpus	Size in tokens	% of corpus
Comedy	427,719	41.2	486,974	44.6
History	296,449	28.5	305,581	28.0
Tragedy	314,527	30.3	298,772	27.4

Table 4. Relative extent of the three broadest play genres within the two corpora.

We see that both ESC:Folio and ESC:Comp consist of roughly 28% histories and 72% tragedies and comedies. (Balancing the genres thus was part of the design of ESC:Comp, but given the need to include plays as wholes or not at all, some minor discrepancy is to be expected; see Demmen, 2020.) Moreover of the two, the *higher* proportion of history plays (by a probably not meaningful 0.5 percentage points) is in ESC:Folio. We can therefore be reasonably confident that the observed lower information density features of Shakespeare's style are *not* merely due to ESC:Folio having less history writing.

In sum, then, if we look just at one corpus versus another, or just at genres, we must always be aware of the possibility of confounding factors. By contrast, examining the full range of variation on a per-text basis allows us to take corpus and genre into account as well as the individual plays. In the next section we explore further approaches to moving beyond aggregate data.

4.2. Contrasts in text-level variability across features, and what they can tell us

In this section we present additional observations made by rendering the spread of the data for the features of grammatical style under analysis in a manner which, compared to those in the previous section, place more emphasis on the individual texts while also making genre immediately evident.

In the figures to be considered, the relative frequency of the given feature in each play is shown by its position, with the shape and fill of its marker indicating its genre and provenance (ESC:Folio or ESC:Comp). Since genre is thus already displayed, we can order the points according to the relative frequency (creating the characteristic sideways-S curve). This layout serves an analysis of the range of variability of a given feature across plays, and of the extent of coincidence of the range in ESC:Folio and ESC:Comp. Figure 2 renders the data for the major part-of-speech features in this way.

This method of visualising the data makes obvious at a glance the distinction between features where the Shakespeare plays and the rest are intermingled over all or nearly all of the range of variability of the feature (e.g. adjectives); features where either Shakespeare or the rest dominates one end of the range of variability (e.g. nouns, dominated at the bottom by Shakespeare and at the top by the rest); and features where either Shakespeare or the rest are more tightly clustered around the central tendency, with the other dominating both extremes (e.g. verbs and adverbs have more of Shakespeare around the central area and less around the top and bottom). These points are, of course, reminiscent of those we observed on the aggregated data – the data considered by-text being the *cause* of aggregate data, it could hardly be otherwise. Similarly, clusters of genre at one or both ends of the sequence can be observed that fit largely with prior observations – such as the predominance of triangular points for histories on the graph for nouns once the 18,000 PMW point is passed.

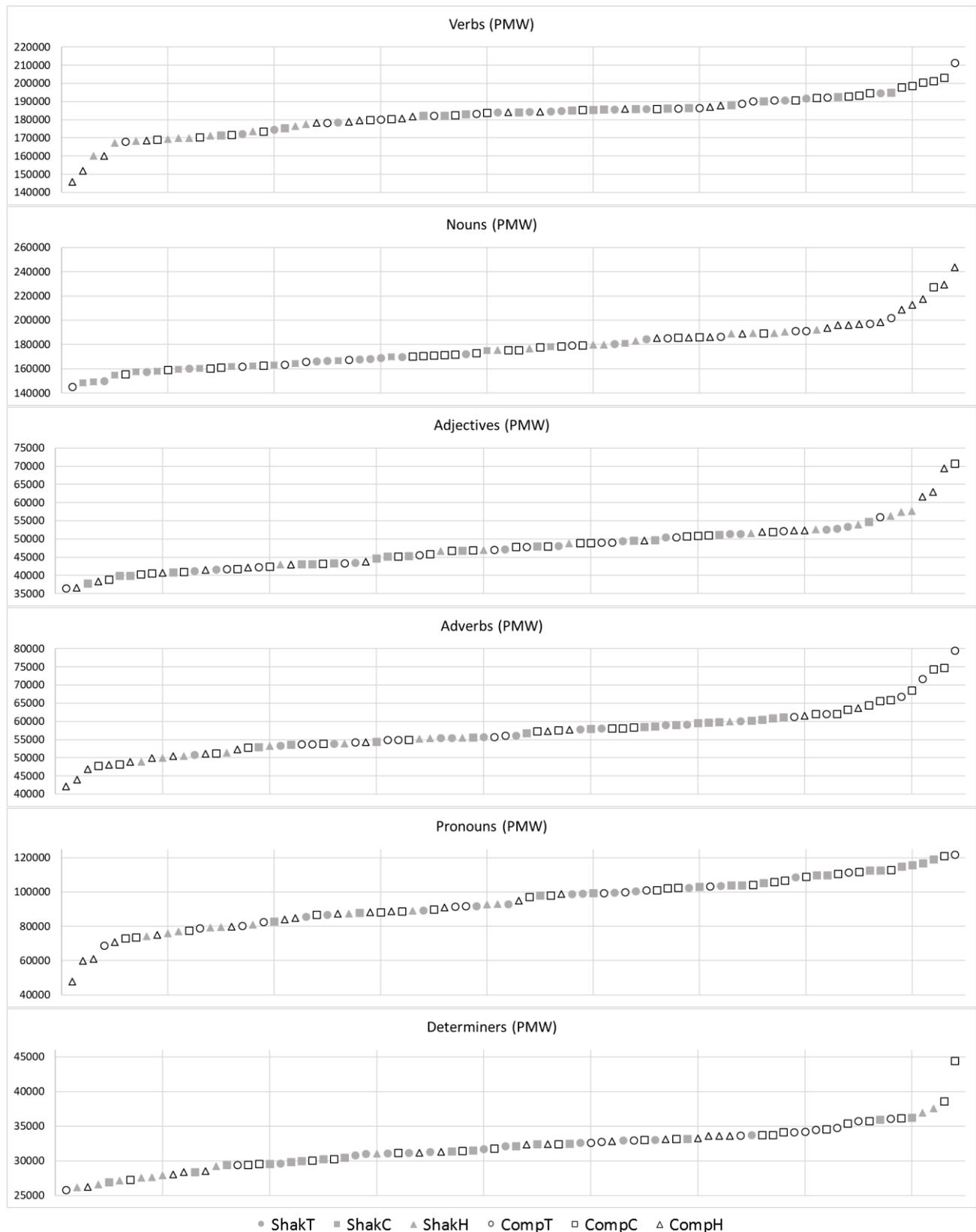


Figure 2. Frequencies of major part-of-speech features by play.

Figures 3, 4, and 5 present in like manner to figure 2 the relative frequencies of the complex grammatical features by play, in three groups: tense and aspect, pronouns, and subordination respectively.

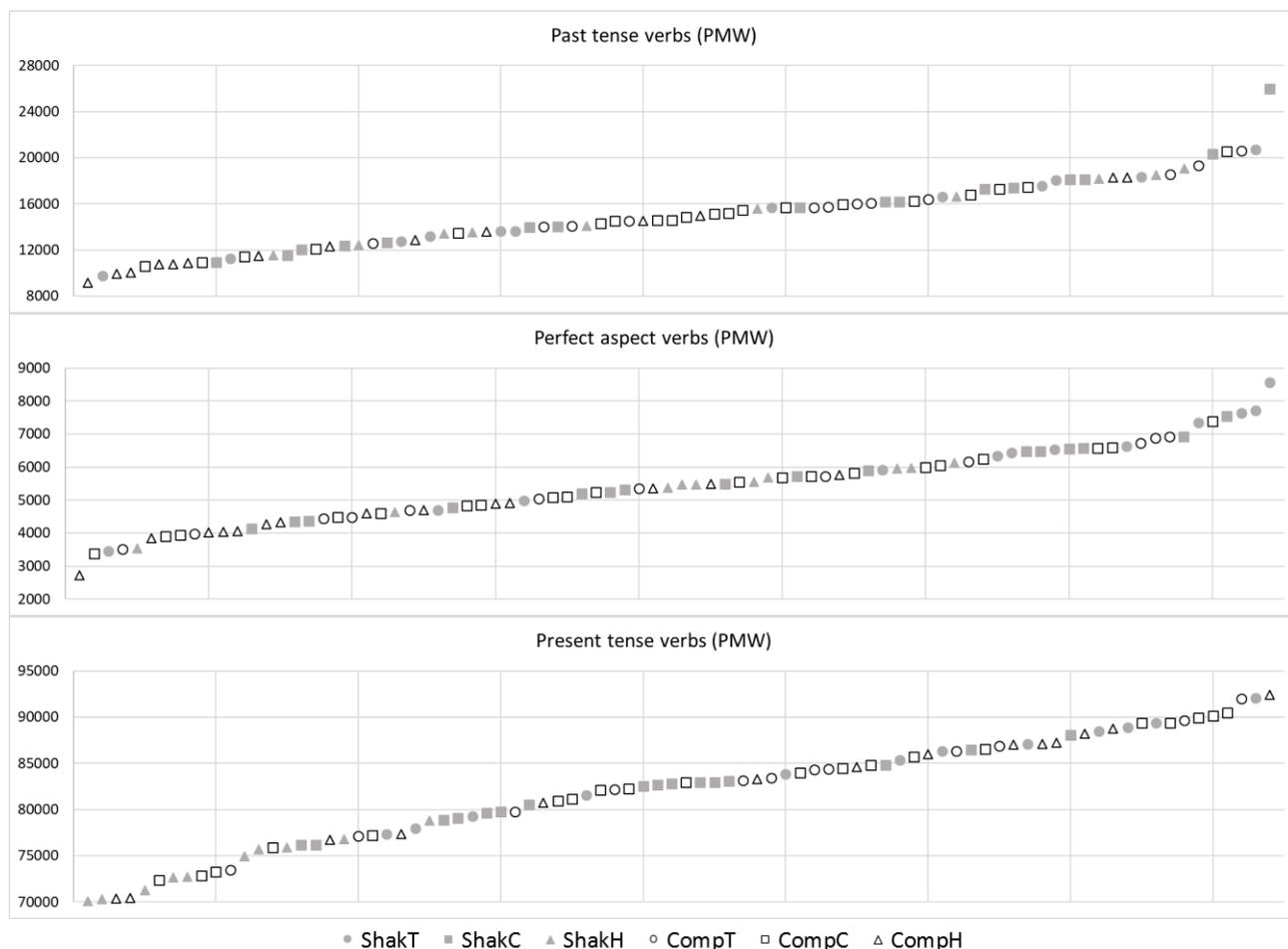


Figure 3. Frequencies of tense/aspect features by play.

We observed in the effect size analysis that frequent use of past tense and perfect aspect distinguishes Shakespeare from the ESC:Comp playwrights; figure 3 shows in detail how this arises. While the full ranges of variability in ESC:Folio and ESC:Comp overlap almost entirely, we see for both past and perfect a concentration of Shakespeare's plays, mostly comedies and tragedies but some histories too, in the upper end of that range. At the lower end we see comparatively few of Shakespeare's plays. This is a clear contrast to the present tense, where Shakespeare and the rest are much more intermingled, although there appears to be a genre effect such that the lower end of the scale is dominated by histories. The observations made previously about narrativity on the basis of these features hold, but we are now able to see that the driving factor is the use of the past and, more clearly, the perfect in many albeit not all of the comedies and tragedies.

Once the plays at the higher end of the scale are identifiable, it becomes straightforward to identify specific loci where past tense, perfect aspect and third person pronouns are thick on the ground, and where the link to narrative function is clear to see, such as this passage in *Coriolanus* act V, scene vi (with the three features indicated as bold, underline, and italic respectively):

AUFIDIUS:

Being banished for it, *he* **came** unto my Hearth,
Presented to my knife *his* Throat: I **took** *him*,
Made *him* joint-servant with me: **Gave** *him* way
 In all *his* own desires: Nay, **let** *him* choose
 Out of my Files, *his* projects, to accomplish
 My best and freshest men, **served** *his* designments

In mine own person: **holp** to reap the Fame
 Which *he* **did** end all *his*; and **took** some pride
 To do myself this wrong: Till at the last
 I **seemed** *his* Follower, not Partner; and
He **wedged** me with *his* Countenance, as if
 I **had been** Mercenary.

CONSPIRATOR:

So *he* **did** my Lord:
 The Army **marvelled** at it, and in the last,
 When *he* **had carried** Rome, and that we **looked**
 For no less Spoil, than Glory.

However, to engage in close analysis of any such example would in the present context be somewhat misleading. The notable finding is neither that the highlighted usages are a particular quality of Shakespeare's dramatic style, nor that passages like the above, where these features cluster thickly, are particular to Shakespeare. Rather the point is that this is a device which Shakespeare's plays exhibit *more often* than the average tendency of the plays in ESC:Comp. Narratively-styled passages such as the above must be deemed a part of the common linguistic toolbox of Early Modern English dramatists; but Shakespeare exhibits a stronger preference for pulling that tool from the box than do his contemporaries.

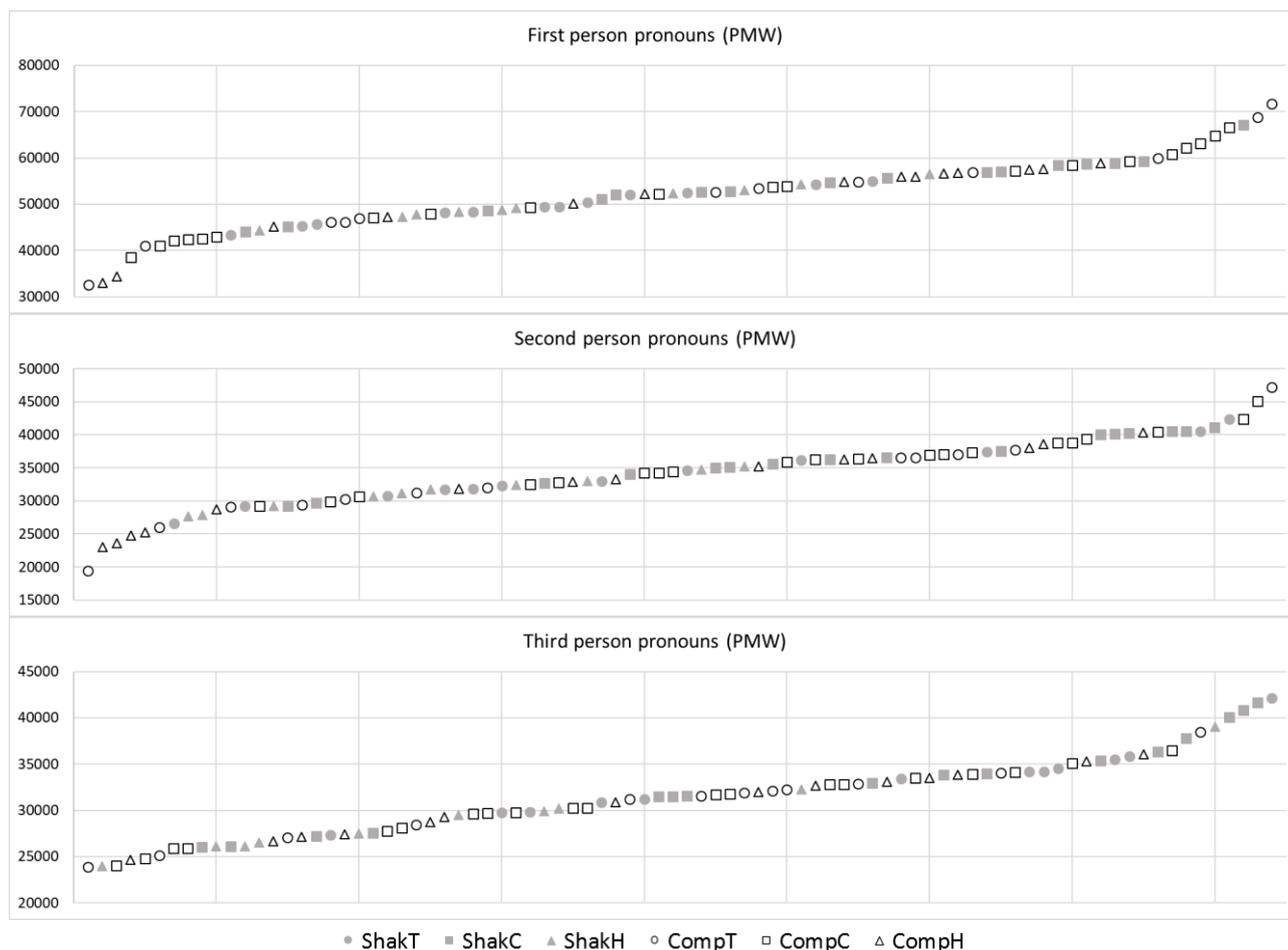


Figure 4. Frequencies of pronoun features by play.

While the virtue of many corpus-based approaches is their capability to generate output data on everything in the corpus – in this case, measurements for every feature in every play – the partial

drawback to this is that some of this data will be inevitably not very interesting. This is well-exemplified for us by the results for first and second pronouns in figure 4. In both, the ESC:Folio plays are somewhat more centrally grouped than the ESC:Comp plays, with both extremities populated solely by non-Shakespeare plays; but, as noted earlier, this is not surprising when one writer is contextualised against many. The third person pronoun data, by contrast, does show us something interesting – but it is the same point on narrativity in some (but not all) comedies/tragedies that emerged above with regard to tense and aspect, and thus needs no further comment.

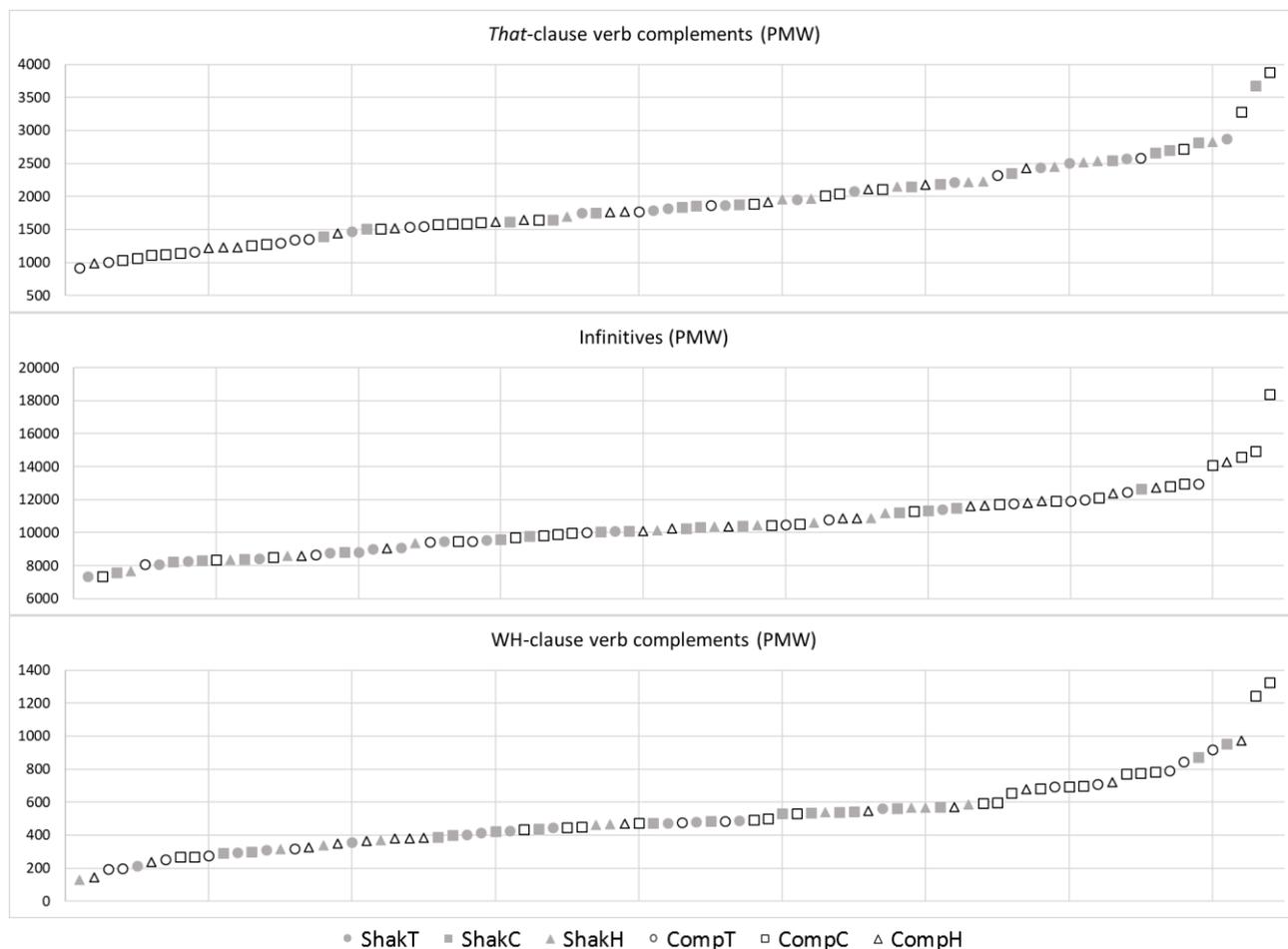


Figure 5. Frequencies of verbal complement subordination by play.

Shakespeare's preference for *that*-clause verb complements was the stand-out result of the earlier effect size analysis, with RRF 1.33. In other words, Shakespeare uses it four times for every three times a playwright in ESC:Comp uses it. This shows up equally strongly in figure 5 at the lower end of the scale. Interestingly, the single highest relative frequency is from ESC:Comp, such that the range of the Shakespeare plays is wholly within the range of the other plays, with much intermingling even to the top of the scale; the overall quantitative effect is, then, driven primarily by the many ESC:Comp plays at the lower end. Genre does not appear to be a major factor in this phenomenon. Meanwhile, the graphs for infinitives and WH-clause verb complements exhibit an all but symmetrically opposite pattern. This is useful confirmation of the former trend (especially because complementiser *that* is homonymous with determiner *that*, such that we must be wary of the impact of part-of-speech errors in ESC:Comp affecting queries such as the search pattern for *that*-clauses). But even collectively these results are insufficient to fully clarify what is going on in terms of subordinate clause complementation; a more detailed analysis, addressing *all* the questions in the paragraph we quoted from Leech and Short (1981: 77) in the introduction to this paper, is both indispensable to this issue and, alas, beyond the scope of the present study.

5. Conclusion

In conclusion, we have identified a number of non-trivial, albeit perhaps not unpredictable, qualities of the grammatical style of Shakespeare relative to a wider sample of early modern playwrights. We arrived at these observations via established concepts and methods in corpus linguistics and stylistics – working at scale on all the plays, without neglecting the variability among the plays, for both Shakespeare and the other playwrights. The methods, both the advanced grammatical corpus queries and the visualisation for analysis of by-text variation, foster a quantitatively informed view of Shakespeare's grammatical style that is largely new to the study of these texts. Of our observations, we would highlight the following in summary:

- Shakespeare's grammatical style prefers features of narrativity including past tense, perfect aspect, and third person pronouns.
- Shakespeare's grammatical style avoids certain phenomena functionally associated with information density (more nouns and determiners), an avoidance which may be seen as a feature of orality, and which is (independent of playwright) associated with comedies/tragedies rather than histories.
- However, first and second person pronouns, also linked to orality, are no more preferred in Shakespeare's grammatical style than in the average style of the other playwrights.

The primary significance of this paper is perhaps not these particular findings, but rather the demonstration that the grammar of Early Modern drama *can* be analysed quantitatively using corpus data in such a way as to throw light on questions of grammatical style. In particular, a window is opened on distinctions between Shakespeare's style and that observable in ESC:Comp that are quantitative rather than absolute. All the grammatical features we looked at were in use in all the plays; what differed was the relative prominence of, or preference for, their use. As we observed in section 2, the most prominent contemporary accounts of Shakespeare's grammar utilise various criteria for what deserves attention (such as features unusual relative to present-day English; or as for Hope, features deemed illicit by prescriptivist scholarship; or as for Blake, topics underserved by modern critical editions of Shakespeare's plays). It is our hope that, in future, such criteria will be tempered by an understanding that attention should *also* be directed to the quantitative criterion of frequency of use relative to the general dramatic style of the time. Both those aspects of Shakespeare's style that are typical of his time, and those where his patterns of linguistic choices, viewed on a large scale, stand out quantitatively, may be relevant for different kinds of stylistic study.

In future work, we hope to address some of the issues which could not be covered within the scope of this paper: to extend the analysis to a much wider range of features (whether Biber's 1988 list, or one derived from Leech and Short, 1981, or some other); to account more comprehensively for distinctions among the dramatic genres (since the comedy/history/tragedy three-way split is only the most general possible schema); and to undertake analysis *in situ* of examples of quantitatively notable features in order to consider the functional load they bear, or the grammatical/discoursal purpose they express. Addressing this last issue via concordance analysis rather than purely quantitatively will allow hypotheses of stylistic function – such as our inference of narrativity and (partially) orality as qualities of Shakespeare's grammatical style that stand out from those of other Early Modern playwrights – to be reconfirmed or refuted. Another specific matter which we consider acutely in need of additional exploration is the different preferences regarding complement subordination between Shakespeare and the rest (see 4.2). Finally, and more speculatively, it may be possible for the corpus-based approaches utilised here to inform future authorship attribution studies, which, as we noted in 3.1, have to date been overwhelmingly based on word forms alone. Critically, future

work can benefit from the example provided here of large-scale analysis which does not neglect the diversity of style exhibited by individual works within the corpus under consideration.

Funding

The research presented in this article was supported by the UK Arts and Humanities Research Council (AHRC), grant reference AH/N002415/1 (*Encyclopedia of Shakespeare's Language*), using technical infrastructure funded by the UK Economic and Social Research Council (ESRC), grant reference ES/R008906/1 (*Centre for Corpus Approaches to Social Science (CASS)*).

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

- Abbott EA (1870) *A Shakespearian Grammar* (3rd edition). London: Macmillan.
- Biber D (1986) Spoken and written textual dimensions in English: Resolving the contradictory findings. *Language* 62: 384-414.
- Biber D (1988) *Variation across Speech and Writing*. Cambridge: Cambridge University Press.
- Biber D, Johansson S, Leech G, Conrad S, and Finegan E (1999) *Longman Grammar of Spoken and Written English*. London: Longman.
- Biese YM (1987) Notes on the use of the ingressive auxiliaries in the works of William Shakespeare. In: Salmon V and Burness E (eds) *A Reader in the Language of Shakespearean Drama*. Amsterdam/Philadelphia: John Benjamins, pp. 329-338.
- Blake N (1983) *Shakespeare's Language: An Introduction*. London: Macmillan.
- Blake N (2002a) *A Grammar of Shakespeare's Language*. Basingstoke: Palgrave.
- Blake N (2002b) Forms of address in Hamlet. In: Lenz K and Möhlig R (eds) *Of dyuersitie & chaunge of language. Essays presented to Manfred Görlach on the occasion of his 65th birthday*. Heidelberg: C. Winter, pp. 305-318.
- Busse B (2006) *Vocative Constructions in the Language of Shakespeare*. Amsterdam: John Benjamins.
- Busse U (1998) Forms of address in Shakespeare's plays: Problems and findings. In: Schulze R (ed) *Making meaningful choices in English. On dimensions, perspectives, methodology and evidence*. Tübingen: Gunter Narr, pp. 33-60.
- Busse U (2002) *Linguistic Variation in the Shakespeare Corpus : Morpho-Syntactic Variability of Second Person Pronouns*. Amsterdam: John Benjamins.
- Calvo C (1992) Pronouns of address and social negotiation in *As You Like It*. *Language and Literature*, 1(1): 5-27.
- Comrie B and Smith N (1977) Lingua descriptive studies: questionnaire. *Lingua* 42(1):1-72.
- Craig H (2018) Authorship, computers, and comparative style. In: Magnusson L and Schalkwyk D (eds) *The Cambridge Companion to Shakespeare's Language*. Cambridge: Cambridge University Press, pp. 168-188. DOI: 10.1017/9781316443668.010.
- Crystal D and Davy D (1969) *Investigating English Style*. London/New York: Routledge.

- Culpeper J, Hardie A, Demmen J, Timperley M and Hughes J (forthcoming) Supporting studying Shakespeare: Structural markup and grammatical annotation enhancing a corpus of the First Folio.
- Cumming G (2014) The New Statistics: Why and How. *Psychological Science* 25(1):7-29.
- Demmen J (2020) Issues and challenges in compiling a corpus of Early Modern English plays for comparison with those of William Shakespeare. *ICAME Journal*, 44: 37-68. DOI: 10.2478/icame-2020-0002.
- Dorst I van (2018) You, thou and thee: A statistical analysis of Shakespeare's use of pronominal address terms. In: Fišer D and Pančur A (eds) *Conference Proceedings of Language Technologies & Digital Humanities 2018*, Ljubljana, Slovenia, 20-21 September 2018, pp. 274-280. Available online at http://www.sdjt.si/wp/wp-content/uploads/2018/09/JTDH-2018_Students_Dorst_You-thou-an-thee-A-statistical-analysis-of-Shakespeare%E2%80%99s-use-of-pronominal-address-terms.pdf
- Egbert J and Biber D (2019) Incorporating text dispersion into keyword analyses. *Corpora* 14(1):77-104.
- Enkvist NE (1973) *Linguistic Stylistics*. The Hague/Paris: Mouton.
- Evert S and Hardie A (2011) Twenty-first century Corpus Workbench: Updating a query architecture for the new millennium. In: *Proceedings of the Corpus Linguistics 2011 conference*, University of Birmingham, UK, 20-22 July 2011. Available online at <https://www.birmingham.ac.uk/documents/college-artslaw/corpus/conference-archives/2011/Paper-153.pdf>
- Fanego T (1990a) Finite complement clauses in Shakespeare's English. Part I. *Studia Neophilologica* 62(1): 3-21.
- Fanego T (1990b) Finite complement clauses in Shakespeare's English. Part II. *Studia Neophilologica* 62(2): 129-149.
- Kakietek P (1987) The perfect auxiliaries in the language of Shakespeare. In: Salmon V and Burness E (eds) *A Reader in the Language of Shakespearean Drama*. Amsterdam/Philadelphia: John Benjamins, pp. 309-317.
- Kermode F (2000) *Shakespeare's language*. London: Allen Lane.
- Hardie A (2007) Part-of-speech ratios in English corpora. *International Journal of Corpus Linguistics* 12(1): 55-81.
- Hardie A (2012) CQPweb – combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics* 17(3): 380-409.
- Hardie A, Hu X and Harvey R (forthcoming) Multidimensional analysis for everybody.
- Hope J (2003) *Shakespeare's Grammar*. London: Arden Shakespeare.
- Hope J (2010) *Shakespeare and language: Reason, eloquence and artifice in the Renaissance*. London: Arden Shakespeare.
- Jeffries L and McIntyre D (2010) *Stylistics*. Cambridge: Cambridge University Press.
- Johnson K (2013) *Shakespeare's English: A Practical Linguistic Guide*. Abingdon/New York: Routledge.
- Johnson K (2019) *Shakespeare's Language: Perspectives Past and Present*. Abingdon/New York: Routledge.

- Jucker AH (2012) 'What's in a name?': Names and terms of address in Shakespeare's *Romeo and Juliet*. In: Chevalier S and Honegger T (eds) *Words, words, words: Philology and beyond. Festschrift for Andreas Fischer on the occasion of his 65th birthday*. Tübingen: Narr Francke Attempto, pp. 77-97.
- Goldberg A (1995) *A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
- Gries STh (2015) The most under-used statistical method in corpus linguistics: multi-level (and mixed-effects) models. *Corpora* 10(1):95-125.
- Ingham R and Ingham M (2011) Subject-verb inversion and Iambic Rhythm in Shakespeare's dramatic verse. In: Ravassat M and Culpeper J (eds) *Stylistics and Shakespeare's language: Transdisciplinary approaches*. London: Continuum, pp. 98-118.
- Jeffries L and McIntyre D (2010) *Stylistics*. Cambridge: Cambridge University Press.
- Leech GN and Short MH (1981) *Style in Fiction: A linguistic introduction to English fictional prose*. London: Longman.
- Mazzon G (2003) Pronouns and nominal address in Shakespearean English: A socio-affective marking system in transition. In: Taavitsainen I and Jucker AH (eds) *Diachronic perspectives on address term systems*. Amsterdam: John Benjamins, pp. 223-249.
- Mulholland J (1967). 'Thou' and 'You' in Shakespeare: a study in the second person pronouns. *English Studies* 48: 34-43.
- O'Halloran K (2007) The subconscious in James Joyce's 'Eveline': a corpus stylistic analysis that chews on the 'Fish hook'. *Language and Literature* 16(3): 227-244. DOI 10.1177/0963947007072847.
- Rayson P, Wilson A and Leech G (2002) Grammatical word class variation within the British National Corpus sampler. In: Peters P, Collins P and Smith A (eds) *New frontiers of corpus research: Papers from the Twenty First International Conference on English Language Research on Computerized Corpora, Sydney 2000*. Amsterdam: Rodopi, pp. 295-306.
- Ravassat M and Culpeper J (eds) (2011) *Stylistics and Shakespeare's language: Transdisciplinary approaches*. London: Continuum.
- Salmon V and Burness E (eds) (1987) *A Reader in the Language of Shakespearean Drama*. Amsterdam/Philadelphia: John Benjamins.
- Shore D (2015) Shakespeare's constructicon. *Shakespeare Quarterly* 66(2): 113-136.
- Short M (1996) *Exploring the Language of Poems, Plays and Prose*. London: Routledge.
- Stein D (2003) Pronominal usage in Shakespeare: Between sociolinguistics and conversational analysis. In: Taavitsainen I and Jucker AH (eds) *Diachronic perspectives on address term systems*. Amsterdam: John Benjamins, pp. 251-307.
- Stockwell P (2020) *Cognitive Poetics: An Introduction* (2nd edition). London: Routledge.
- Stubbs M (2001) *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell.
- Taylor EW (1087) Use of *eth* and *es* endings of verbs in the First Folio. In: Salmon V and Burness E (eds) *A Reader in the Language of Shakespearean Drama*. Amsterdam/Philadelphia: John Benjamins, pp. 349-369.
- Wales K (1983) *Thou* and *you* in Early Modern English: Brown and Gilman re-appraised. *Studia Linguistica* 37(2): 107-125.

Wales K (1985) Generic 'your' and Jacobean drama: The rise and fall of a pronominal usage. *English Studies* 66(1): 7-24.

Wales K (1987) An aspect of Shakespeare's dynamic language: A note on the interpretation of King Lear III. VII.113: 'He childed as I father'd!' In: Salmon V and Burness E (eds) *A Reader in the Language of Shakespearean Drama*. Amsterdam/Philadelphia: John Benjamins, pp.181-192.

Bionotes

Andrew Hardie is a Reader in Linguistics, specialising in corpus linguistics. He is the author, with Tony McEnery, of the book *Corpus Linguistics: Method, Theory and Practice* (2012), and is also known for creating the CQPweb corpus analysis software. He is the technical lead of the *Encyclopedia of Shakespeare's Language* project.

Isolde van Dorst is a doctoral researcher. Her PhD focuses on politeness in English business communication. In addition, she has carried out research on Shakespeare's use of second person pronominal address terms and is the senior entry compiler on the *Encyclopedia of Shakespeare's Language* project.

¹ Throughout, we consider *style* from the perspective of authorial style rather than genre style, text style, etc.

² *Variability* here is approximately equivalent to what is, in corpus linguistics, usually dubbed (*text*) *dispersion*. We avoid framing the present investigation in those terms, however, because dispersion is currently an issue under some debate (e.g. see Egbert and Biber, 2019), engaging with which lies outside this paper's scope.

³ Most of the work reviewed does not explicitly exclude Shakespeare's non-dramatic poetry from consideration. However, discussion is invariably focused on the plays.

⁴ Shore presents the forms of these constructions in CQPweb's Simple Query syntax, formally the *Common Elementary Query Language* (CEQL), invented by Stefan Evert (Evert and Hardie, 2011: 14-15). To avoid CEQL formalism, we schematise the constructions informally here.

⁵ These figures are drawn from ESC:Folio, see 3.2.

⁶ <http://wp.lancs.ac.uk/shakespearelang/>

⁷ <http://ucrel.lancs.ac.uk/claws>

⁸ The full CQP tutorial is available at <http://cwb.sourceforge.net/documentation.php#tutorials>

⁹ <http://pcre.org/>

¹⁰ <http://ucrel.lancs.ac.uk/claws6tags.html>

¹¹ The play codes in ESC:Folio, shown in Figure 1, are the Arden edition abbreviations of Shakespeare's titles. The more numerous play codes in ESC:Comp are omitted from figure 1 for clarity.

¹² The need for interpretation of quantitative data to be tempered by detailed understanding of the nature of the content of the corpus or corpora is a commonplace of corpus linguistic methodology.