CHAPTER 21

GENRE

BRETT GREATLEY-HIRSCH AND SARAH NEVILLE

Though the 1623 First Folio apportions Shakespeare's dramatic oeuvre into three seemingly self-evident classes of comedy, history, and tragedy, during Shakespeare's lifetime the demarcation of these three genres was far from inevitable. In his 1598 *Palladis Tamia* or *Wit's Treasury*, Francis Meres had separated Shakespeare's plays into only two categories, comedy and tragedy:

> As *Plautus* and *Seneca* are accounted best for Comedy and Tragedy among the Latines: so *Shakespeare* among [the] English is the most excellent in both kinds for the stage; for Comedy, witnes his *Ge*[*n*]*tlme*[*n*] *of Verona*, his *Errors*, his *Loue labors lost*, his *Loue labours wonne*, his *Midsummers night dreame*, & his *Merchant of Venice*: for Tragedy his *Richard the 2. Richard the 3. Henry the 4. King Iohn*, *Titus Andronicus* and his *Romeo* and *Iuliet*.[1]

Conspicuously absent from Meres's list is the third genre that the publishers of the First Folio sought to highlight on its title page, the English history play, a form that critics associate with Shakespeare's rise in the commercial theatre.[2] Instead, Meres subsumes history plays under the generic label of 'Tragedy' and he either dismisses or does not associate the collaboratively written *Henry VI* plays, among the earliest of Shakespeare's works, with Shakespeare at all. Together, these details suggest that as late as 1598, the genre of 'history' as a legible dramatic mode had not yet been sufficiently codified in the English aesthetic imagination for Meres to draw readers' attention to the form. By 1623, however, the word 'history' could so legibly function to distinguish a specific type of play that it could serve to advertise a large and expensive volume.

[1] Francis Meres, *Palladis Tamia. Wits Treasury Being the Second part of Wits Common wealth* (London, 1598; STC 17834), 2O2r.

[2] See Paulina Kewes 'The Elizabethan History Play: A True Genre?', in *A Companion to Shakespeare's Works*, Vol. 2, eds., Richard Dutton and Jean E. Howard, 4 vols (Malden: Blackwell, 2003), 170–193.

The evidence of both *Palladis Tamia* and the First Folio title page show that dramatic genres functioned to prescribe interpretation of Shakespeare's texts by shaping them to fit existing forms. As a vocabulary shared between authors and audiences, the rhetoric of genre allowed Meres to fit Shakespeare into a large and diverse community of classical, medieval, and contemporary writers and it offered the First Folio publishers a way of distinguishing their wares in the marketplace of books in early modern London.[3] But while this external evidence offers us some sense of the attitudes towards dramatic genre held by Shakespeare's contemporaries, it reveals little of Shakespeare's own attitudes towards dramatic form.

Turning to a discussion of genre from within Shakespeare's plays allows a slightly different picture to emerge. In the second quarto text of *Hamlet*, Shakespeare appeared to assume that London audiences could appreciate the existence of several distinct dramatic genres well enough that he could use them for metatheatrical stage humour. In introducing the players, Polonius remarks that they are so famously

> The best actors in the world, either for tragedy, comedy, history, pastoral, pastoral-comical, historical-pastoral, scene individable, or poem unlimited. Seneca cannot be too heavy, nor Plautus too light. For the law of writ, and the liberty, these are the only men. (7.310–313)

Critics have generally located the humour of the speech in the adjectival fusions 'pastoral-comical' and 'historical-comical', seeing them less as authentic early modern genres than as evidence of Polonius's growing rhetorical evasion. The counsellor who earlier wishes not to 'crack the wind of [a] poor phrase' (3.107) has begun to lose track of his language: 'what was I about to say?' (6.49). As the play progresses, Polonius becomes increasingly unable to distinguish between matter and art, and his adjectival excesses are part of this larger characterization. The speech in the revised 1623 F1 text of *Hamlet* even goes further to supply new hybrids to the generic catalogue, adding 'tragical-historical' and 'tragical-comical-historical-pastoral', a capacious but incongruous category, to Polonius's bewildered and bewildering list.[4]

These adjectival flourishes from the First Folio are so attractive that editors of *Hamlet*, who otherwise follow the second quarto, generally emend to add them, thereby furthering readings that see generic litany as yet another characteristic of Polonius's affectation. For example, in finding the list 'pompous', David Bevington's note on the passage adopts Hamlet's irritation with Polonius, amplifying critics' traditional sympathies with the character of the prince.[5] Similarly, Philip Edwards claims that in

[3] On the publication of the First Folio, see Ben Higgins, *Shakespeare's Syndicate: The First Folio, its Publishers, and the Early Modern Book Trade* (Oxford: Oxford University Press, 2022).

[4] On the revision of *Hamlet*, see Richard Dutton, *Shakespeare, Court Dramatist* (Oxford: Oxford University Press, 2016).

[5] David Bevington, ed., *The Complete Works of Shakespeare*, sixth edition (New York: Pearson, 2008), 1116.

Polonius 'Shakespeare illustrates the inadequacy of categorizing the modern drama of his day by the old genres.'[6] As Ian Smith has noted, 'critics identify with Hamlet personally and see in his doubt and skepticism a philosophy that resonates with their world view and towards a modern social ethos';[7] if so, the extended catalogue of preposterously conflated genres is a useful mechanism for allowing editors and commentators to share in Hamlet's mockery of Polonius as an old fool who fundamentally misunderstands how drama works.

But though the speech on genre is often played for laughs, it is unclear whether Polonius himself is responsible for the contents of his list, his obsessive classifications indicative of an equivocal mind, or whether Shakespeare was using Polonius's catalogue to meta-theatrically deride the hodgepodge of dramatic modes that appeared on the early modern stage. In a performance note on the passage in the *New Oxford Shakespeare*, Terri Bourus observes that 'Polonius sometimes reads from a playbill', a staging that divests the famously pedantic counsellor from responsibility for the catalogue and which, as the list continues, allows him to become increasingly enthusiastic, incredulous, or exasperated with the players' absurd self-advertising. Polonius's account of the players' genres finds its basis in the vocabularies advertising Renaissance printed playbooks, revealing, even despite Shakespeare's supposed indifference to print publication, his sophisticated awareness of how plays were marketed in print.[8] The external evidence of title pages thus confirms what Harold Jenkins astutely remarked in his commentary on the passage: 'what Polonius begins as a sober, if inappropriate, catalogue ends in self-parody—through which Shakespeare no doubt smiles at contemporary dramatic olios'. What Jenkins seems not to have realized, however, is that the olios, or heterogeneity, of early modern dramatic genres was as much a commercial feature of the paratexts of printed plays as it was a function of plays' essential internal form. In other words, the ways that Renaissance plays were sold to readers later determined how critics learned to read them.

External Evidence: Playbooks and Paratexts

The title pages of playbooks were primary locations where booksellers could market plays to would-be reading customers by drawing their attention to authorial, social, or

[6] Philip Edwards, ed., *Hamlet, Prince of Denmark*, updated edition (Cambridge: Cambridge University Press, 2003), 147.

[7] Ian Smith, *Black Shakespeare: Reading and Misreading Race* (Cambridge: Cambridge University Press, 2022), 158–159.

[8] Alan B. Farmer, 'Shakespeare as Leading Playwright in Print, 1598–1608/09', in *Shakespeare and Textual Studies*, eds. Margaret Jane Kidnie and Sonia Massai (Cambridge: Cambridge University Press, 2015), 87–104.

thematic features, including genre, signalling, for example, to a London book-buyer that Polonius's 'heavy' Seneca wrote tragedies and 'light' Plautus penned comedies.[9] By this measure, *Hamlet* itself seems not to suffer from generic confusion, at least, not in the 1623 Folio. In its 'Catalogue of the seuerall Comedies, Histories, and Tragedies contained in this Volume', the First Folio lists *Hamlet* seventh among the tragedies, sandwiched between *Macbeth* and *King Lear*. Less appreciated, however, is the fact that Shakespeare's *Hamlet* was not identified as a tragedy in its earliest incarnations in print and, ironically, the earliest generic classification of *Hamlet* in its first two editions mirrors Polonius's incongruous and bewildering language in the First Folio. For Nicholas Ling and John Trundle, the earliest publishers seeking to advertise the play to readers, *Hamlet* was best understood not as a tragedy, but as a 'Tragicall Historie', the first of only two plays in the entire period to be so identified.[10]

While in performance a phrase like 'tragical history' is often played for laughs, generic mixing makes more sense when removed from the dramatic context of Polonius's stage to appear on a stationer's page. In the competitive environment of the London book trade, Ling and Trundle's decision to market *Hamlet* in 1603 (STC 22275) and again in 1604 (STC 22276) specifically as a 'Tragicall Historie' was a deliberate and strategic choice designed to exploit the growing popularity of history plays in print.[11] Between 1578 and 1602, two dozen other playbooks had been advertised as histories, with two-thirds of those printed in the last five years of that span.[12] For a speculative publisher like Ling looking to maximize the amount of profit from his publication of Shakespeare's *Hamlet*, adopting the generic category of history on the title page might have seemed like a clever idea, particularly if he got the best of both words by modifying the noun 'tragedy' to form the adjective 'tragical', as used on the title pages of popular plays by Christopher Marlowe.[13] Other stationers who had recently published Shakespeare's plays advertised the genres of comedy and tragedy with the modifying adjectives of 'conceited' or 'excellent', and perhaps Ling, who had never before published one of Shakespeare's plays, was strategically maximizing the marketing potential available on his title page.[14] But while an adjective like 'tragical' was useful for affiliating *Hamlet* with

[9] Alan B. Farmer and Zachary Lesser, '"Vile Arts": The Marketing of English Printed Drama, 1512–1660', *Research Opportunities in Renaissance Drama*, 39 (2000), 77–165. All but one of the twenty Renaissance editions of Seneca's plays were identified on their title page as tragedies, while the sole edition of Plautus, William Warner's translation of *Menaechmi*, was sold as 'A pleasant and fine Conceited Comaedie' (STC 20002).

[10] The other is Christopher Marlowe's *Doctor Faustus*, first printed in 1604 (STC 17429).

[11] On Ling, see Terri Bourus, *Young Shakespeare's Young Hamlet: Print, Piracy, and Performance* (London: Palgrave, 2014).

[12] The claims of this and the following two paragraphs are indebted to Alan B. Farmer and Zachary Lesser, ed, *DEEP: Database of Early English Playbooks*: deep.sas.upenn.edu.

[13] With the exception of *Massacre at Paris* (1594), all of Marlowe's printed playbooks prior to *Faustus* identified the play as a tragedy or featured the word 'tragical' on their title pages.

[14] Q1 *Romeo and Juliet* was advertised as an 'Excellent Conceited Tragedie', but when Cuthburt Burby reprinted it in 1599, he replaced 'conceited' with 'lamentable'. Yet Burby liked 'conceited' enough as a marker of wordplay to *Love's Labour's Lost* (1598, STC 22294).

the published works of Marlowe, it was the noun 'history' which was of particular value in advertising what would become Ling's 1603 and 1604 editions of the play. Another stationer, Andrew Wise, had had enormous success with *The Historie of Henry the Fourth*, first published in 1598 (STC 22279a) and then immediately reprinted twice (STC 22280, 22281). To attentive booksellers of playbooks at the turn of the century like Ling, Wise's three editions of *Henry IV* indicated that a history play by Shakespeare could have tremendous popularity with London readers.

Wise could serve as a model for Ling because he was an especially savvy stationer who well understood the value of a title page as space for advertising copy. With Cuthbert Burby, Wise was among the first stationers to add Shakespeare's name to title pages, and he also held the rights to Shakespeare's *Richard II* and *Richard III*, playbook titles that, judging by their quick reprint rates, were exceptionally popular with the reading public. But though the First Folio would later rebrand plays about English monarchs like *Richard II* and *Richard III* as histories, Wise had first chosen to advertise them as tragedies, a successful generic brand for these quartos that remained even after Wise transferred rights to publish them to Matthew Law in 1603.

Wise's success marketing Shakespeare in print in two discrete genres, tragedy and history, provides a context for Ling and Trundle's use of the phrase 'Tragicall Historie' to advertise *Hamlet* in 1603 and 1604. Ling and Trundle used genre to brand their playbooks in accordance with a brief trend for publishing history plays, and they were not the only stationers to do so. In 1600, James Roberts branded his quarto edition of *The Merchant of Venice* (STC 22296) as 'the most excellent Historie', a term that did not appear in the Stationers' Registers entry for the play on 22 July 1598, indicating that Roberts had purposefully added the word.[15] In the five years prior to Ling and Trundle composing the title page of *Hamlet* in 1603, ten plays advertised on their title pages as histories had appeared in London's bookstalls—and three of them were by Shakespeare.[16] At the turn of the century, printed history plays were in fashion, influencing how Shakespeare's *Hamlet* was initially marketed to London readers—and it may have caught Shakespeare's attention enough that he had Polonius take note of the trend.

But trends fade, and as the First Folio catalogue testifies, *Hamlet* did not last long as a 'tragicall historie'. Ling transferred his rights to the play to stationer John Smethwick in 1607, and when it came time for Smethwick to reprint the *Hamlet* third quarto in 1611, he was less convinced that the hybrid category that had previously sold the play was still of use. Smethwick divested the trendy term 'historie' entirely, rebranding the play simply as 'THE TRAGEDY OF HAMLET, Prince of Denmarke' (STC 22277).[17] Smethwick was a member of the 1623 Folio syndicate, and his attention to generic rebranding suggests that he inclined

[15] See Edward Arber, ed., *A Transcript of the Registers of the Company of Stationers of London; 1554–1640 A.D.*, 5 vols (London: Printed Privately, 1875), 3.122 (Register C, fol. 39v).

[16] The first two editions of *The Taming of a Shrew* (STC 23667, 23668), sometimes attributed to Shakespeare and dated 1594 and 1596, also identified that play on the title pages as a 'pleasant conceited historie'.

[17] Later, in 1631, Smethwick would also alter the title of another quarto 'historie' that he obtained from Ling along with *Hamlet*, correcting the genre of that play to better accord with contemporary categories,

towards, if not instigated, the Folio's careful streamlining of Shakespeare's plays into the generic triad of comedies, histories, and tragedies that has since become canonical.

The dramatic fiction that Polonius was created to serve may have prevented his silly speech about genre from being treated seriously by critics, but the evidence of title pages reveals that the old counsellor had an accurate read on the literary culture in which Shakespeare wrote. The varied adjectives on playbook title pages testify that, for late Elizabethan and early Jacobean stationers, and for *Hamlet*'s players, the 'hodgepodginess' of genre was a useful marketing ploy. Whoever is responsible for the generic categories that Polonius utters to such amusing effect, the effect is to suggest that the players, themselves rendered notoriously unfashionable by the 'humor of children', are indeed so outrageously dated that they characterize the plays they perform with a crushing wealth of adjectives.

Polonius's catalogue suggests that genre could at times be a matter of marketing. For instance, stationers sometimes used adjectives to modify generic nouns on playbooks published before 1600, allowing hybrids like 'tragical history' to be briefly in vogue before later settling into more familiar, singular forms. This kind of fluidity in the understanding of play genres has not always been sufficiently appreciated by later scholars. As Andy Kesson has shown in his study of John Lyly, the genres ascribed to plays by stationers can obscure the more flexible generic realities of authorial composition, and he notes that 'comedy, history and tragedy are not comprehensive, self-evident or discrete as descriptors of early modern playwriting'.[18] To put it slightly differently, what eventually becomes *the* official generic triad for Shakespeare's plays in the First Folio was less the product of Shakespeare's authorship and more the result of the decisions of stationers who needed to figure out how to advertise and arrange the contents of a large tome.

When presented with a comprehensive corpus of Shakespeare's dramatic works, the publishers of the First Folio organized the plays according to a deceptively simple logic that has appeared straightforward and self-evident to later readers. But the evidence of the book trade shows that generic labels were fluid and contingent during Shakespeare's lifetime, and this flexibility allowed the publishers of the First Folio to alter the generic labels of many of Shakespeare's plays that had previously appeared in print. The stationers changed the 'pleasant conceited historie' of *The Taming of the Shrew* into a comedy, altered 'The Tragedie of King Richard the Second' into a history, and transformed the 'true chronicle historie' of *King Lear* into a tragedy. Yet it is unclear why, exactly, they recategorized the genres of previously printed plays. Were they simply responding to the demands of the playbook market? Or did the stationers recognize something essential in the texts themselves that merited reading the plays in this way? And can we tell the difference? Kesson's work reveals that, because literary critics have been predisposed to read genre through paratextual information, they have sometimes misread authors' own sense of genre. Because we have always read

and rechristening it as 'A WITTIE AND PLEASANT COMEDIE Called *The Taming of the Shrew*' (STC 22327).

[18] Andy Kesson, 'Was Comedy a Genre in English Early Modern Drama?', *British Journal of Aesthetics*, 54 (2014), 224.

Shakespeare's plays through the lens of the First Folio categories, it turns out we know very little about whether Shakespeare's plays sort into comedies, histories, and tragedies without the invisible hand of a fiscally motivated third party.

Internal Evidence: What's Genre to Shakespeare, or Shakespeare to Genre?

> Can one identify a work of art, of whatever sort, but especially a work of discursive art, if it does not bear the mark of a genre, if it does not signal or mention it or make it remarkable in any way?[19]

Computational methods offer a way to read outside of the paratextual categories imposed upon plays by booksellers, relying entirely upon the internal evidence of the plays themselves to derive a sense of how Shakespeare used generic form. By using the methods of computational stylistics to analyse the thirty-six plays categorized by the First Folio as comedies, histories, and tragedies, we seek to identify latent patterns in language use across the corpus, observe linguistic affinities and differences between the plays, and thereby determine whether and to what extent genre is a constraining factor on Shakespeare's dramatic writing.

We first construct a corpus of the thirty-six plays contained in the First Folio, using the Folger Digital Texts (folgerdigitaltexts.org) as base texts for our analysis. The Folger Digital Texts were chosen because they are suitably formatted for natural language processing with annotations defining structure and marking parts of speech (allowing differentiation, say, between the word *hail* used as a singular noun, verbal interjection, or verbal infinitive, and between words used in dialogue, stage directions, etc). Further, because the Folger Digital Texts are open-access, they are available to readers who wish to verify our findings. These are, of course, critically edited eclectic texts incorporating readings from editions beyond the First Folio, but because the Folger texts were all edited and modernized by Barbara A. Mowat and Paul Werstine, they offer methodological consistency and avoid the thorny issue of historical spelling variation.[20] Table 21.1 lists the plays in our corpus, their generic First Folio designations and total words of dialogue counted by *type* (unique word-forms) and *token* (instances of those word-forms).

With the corpus so constructed, we can use software to count any quantifiable feature of the texts and perform algorithmic procedures. To get a basic overview of the

[19] Jacques Derrida, 'The Law of Genre', *Critical Inquiry*, 7 (1980), 64.

[20] See Anupam Basu, '"Ill shapen sounds, and false orthography": A Computational Approach to Early English Orthographical Variation', in Laura Estill, Diane K. Jakacki, and Michael Ulyot, ed., *Early Modern Studies after the Digital Turn* (Toronto: Iter Press, 2016), 167–200.

Table 21.1 First Folio plays in the corpus

Title	F1 genre	Types	Tokens
All's Well That Ends Well	Comedy	3,870	22,863
Antony and Cleopatra	Tragedy	4,340	24,022
As You Like It	Comedy	3,591	21,214
The Comedy of Errors	Comedy	2,815	14,474
Coriolanus	Tragedy	4,436	26,724
Cymbeline	Tragedy	4,736	27,028
Hamlet	Tragedy	5,177	29,903
1 Henry IV	History	4,250	24,107
2 Henry IV	History	4,569	25,899
Henry V	History	5,048	25,626
1 Henry VI	History	4,188	20,316
2 Henry VI	History	4,521	24,560
3 Henry VI	History	4,018	23,430
Henry VIII	History	3,986	23,531
Julius Caesar	Tragedy	3,210	19,097
King John	History	3,976	20,445
King Lear	Tragedy	4,622	25,376
Love's Labour's Lost	Comedy	4,155	20,891
Macbeth	Tragedy	3,613	16,582
Measure for Measure	Comedy	3,653	21,259
The Merchant of Venice	Comedy	3,647	21,003
The Merry Wives of Windsor	Comedy	3,576	21,282
A Midsummer Night's Dream	Comedy	3,317	16,140
Much Ado About Nothing	Comedy	3,330	20,942
Othello	Tragedy	4,193	26,201
Richard II	History	4,064	21,498
Richard III	History	4,579	28,295
Romeo and Juliet	Tragedy	4,162	24,052
The Taming of the Shrew	Comedy	3,596	20,442
The Tempest	Comedy	3,486	16,268
Timon of Athens	Tragedy	3,626	17,985
Titus Andronicus	Tragedy	3,778	19,834
Troilus and Cressida	Tragedy	4,715	25,648
Twelfth Night	Comedy	3,376	19,584
The Two Gentlemen of Verona	Comedy	3,039	17,068
The Winter's Tale	Comedy	4,300	24,808

play data, we begin by calculating summary statistics. By total dialogue, *Hamlet* is the longest play with 29,903 tokens, followed by *Richard III* (28,295) and *Cymbeline* (27,028); the shortest is *The Comedy of Errors* (14,474 tokens), followed by *A Midsummer Night's Dream* (16,140) and *The Tempest* (16,268). We can visualize the summary statistics as a violin plot (Figure 21.1): the filled 'violin' shapes reflect the distribution of the plays, the white dots mark the mean averages, and the white lines the standard deviations above and below the means. As we can see in Figure 21.1, First Folio comedies are typically shorter in total dialogue (with a mean of 19,874 tokens) than histories and tragedies (with means of 23,771 and 23,538 respectively). Tragedies show the greatest variation in

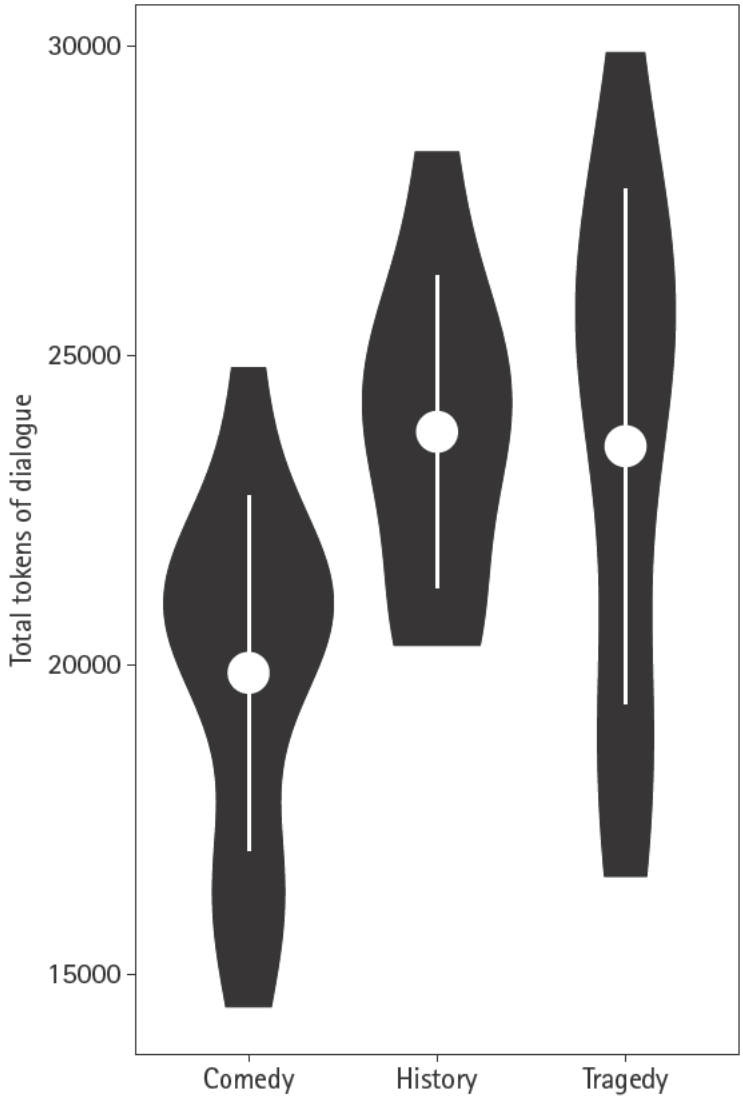


FIGURE 21.1 Violin plot of First Folio genres by total tokens of dialogue.

total dialogue with a larger standard deviation (4,166 tokens) than comedies (2,866) and histories (2,528); thus, while histories cannot boast the longest or shortest plays among their number, they are more consistent when it comes to total dialogue.

Empirical findings like these allow us to situate individual plays or discrete genres in their broader contexts, to make informed comparisons, and to draw principled generalizations. There are, admittedly, limitations to the kinds of statistical inferences we can make, some of which derive from the data sample (e.g. the unavoidable inclusion of texts thought to be cut for performance, such as *Macbeth*) and others from the features selected for analysis (e.g. quantity of dialogue does not, in and of itself, directly map onto stage-time, since it conveys nothing about verbal delivery and other stage business). Some broad patterns emerge (i.e. Shakespeare typically wrote comedies with shorter total dialogue than histories and tragedies), but the amount of overlap renders length an unreliable stylistic discriminator of genre. We would, for instance, risk misclassifying *Julius Caesar* on this basis: at 19,097 tokens, it is unusually shorter than other tragedies but within the normal range for comedy.

Counted in this way, genre appears to have had limited impact on the volume or length of Shakespeare's dramatic dialogue. Could genre instead have constrained Shakespeare's vocabulary, such that certain genres necessitated painting with a broader linguistic palette than others? One standard statistical measure of vocabulary 'richness' is the type–token ratio, calculated by dividing the total number of types (unique word-forms) by the total tokens (instances of those word-forms) in a text: the higher the type–token ratio, the more diverse a text's vocabulary. This simplicity comes at a cost: the type–token ratio treats all words as if they were equivalent, but natural language does not work this way—meaningful communication means different words are used at differing rates, depending on context. An alternative solution is to use Shannon Entropy, a more nuanced statistical measure of vocabulary 'complexity' based on probabilities of word distribution.[21] Although the results will vary depending on the kinds of words used, a rule of thumb is that a text containing a large number of types represented by few tokens (e.g. a dramatis personae in which every name is exclusive) will often have a higher Shannon Entropy rate than a text of the same length containing fewer types distributed across many tokens (e.g. a legal agreement in which the same terms are repeated ad nauseum to avoid ambiguity).

To make valid comparisons between texts of different length, we calculate Shannon Entropy rates for each of the Folio plays using only the first 10,000 words of dialogue (see Figure 21.2). *1 Henry VI* (6.503) and *Henry V* (6.447) have the highest Shannon Entropy rates, and *Much Ado* (6.100) and *Merry Wives* (6.114) the lowest. With a generic mean of 6.246, comedies typically have lower Shannon Entropy rates than tragedies (mean 6.333) and histories (mean 6.371). These results suggest that, in terms of vocabulary 'complexity', plays categorized as comedies in the First Folio are generally more linguistically

[21] C. E. Shannon, 'A Mathematical Theory of Communication', *Bell System Technical Journal*, 27 (1948), 379–423, and 'Prediction and Entropy of Printed English', *Bell System Technical Journal*, 30 (1951), 50–64. On its use in stylistic analysis, see Hugh Craig and Brett Greatley-Hirsch, *Style, Computers, and Early Modern Drama: Beyond Authorship* (Cambridge: Cambridge University Press, 2017).

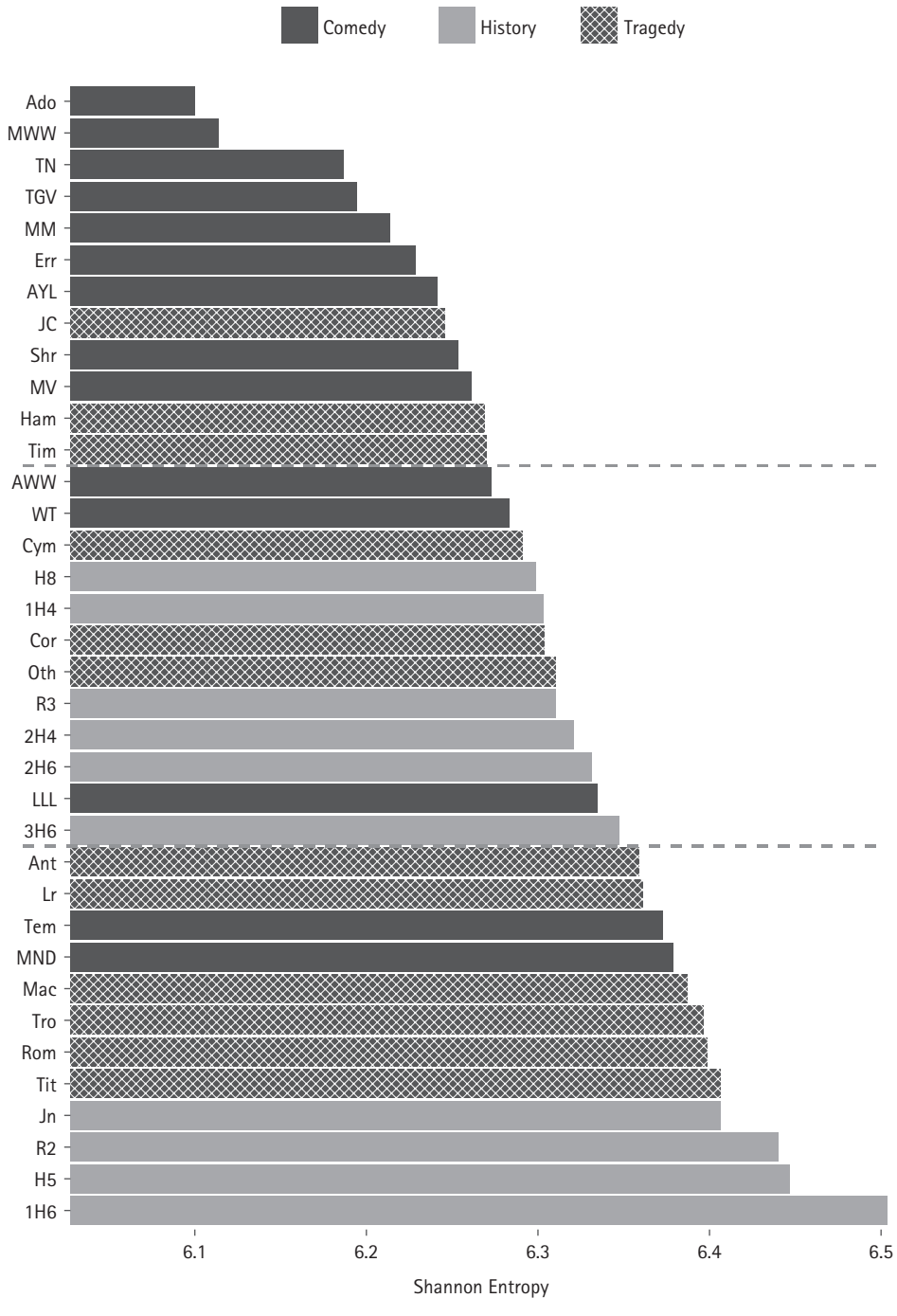


FIGURE 21.2 Shannon Entropy for First Folio plays using the first 10,000 tokens of dialogue.

predictable than those classed as histories or tragedies. In some ways, this is an unsurprising result: comedies often rely on repetition for their humour and are structured to place more emphasis on stichomythic interactions between characters than histories or tragedies, which, as our first test showed, typically involve larger vocabularies. There are exceptions, of course: on this measure, *A Midsummer Night's Dream* (6.379) and *The Tempest* (6.373) have much higher Shannon Entropy rates than other comedies, which predominate the first tercile; conversely, *Julius Caesar* (6.245), *Hamlet* (6.269), and *Timon of Athens* (6.270) are the lowest-scoring Folio tragedies, which otherwise predominate the third tercile.

Other techniques allow us to explore Shakespeare's dramatic vocabulary in finer detail. To test whether the First Folio genres use distinctive vocabulary, we employ Zeta, a method more commonly used in attribution studies to classify texts of unknown authorship. The method works by selecting 'marker' words that distinguish texts in a reference set (e.g. those designated by the First Folio as histories) from those in a comparison set (e.g. comedies and tragedies combined, or 'non-histories').[22] Zeta is concerned with word distribution, and so this selection process is accomplished by calculating index scores for every word (expressed as the proportion of texts in the reference set where the word is present added to the proportion of texts in the comparison set where the word is absent, and then vice versa) and then ranking them.

Taking the First Folio histories as our reference set and non-histories as our comparison, a word present in all ten histories and absent from all twenty-six non-histories would receive an index of 2 (=10/10 + 26/26), the highest score possible; a word present in five histories and absent from thirteen non-histories would receive an index of 1 (=5/10 + 13/26), a neutral score, as would a word present in all plays of both sets (=10/10 + 0/26); and a word absent from all histories and present in all non-histories would receive an index of 0 (=0/10 + 0/26), the lowest possible score. The higher a word's index score, the greater its positive discriminating power; for Zeta, the strength of a marker word is a matter of it being consistently present and correspondingly absent within the respective sets, and not how frequently it appears in any text.

We use Zeta to rank words of dialogue distinguishing First Folio comedies from non-comedies, histories from non-histories, and tragedies from non-tragedies, selecting those with an index score of 1.5 or above as markers. Table 21.2 summarizes the results, listing the marker words (with NUPOS v2 part-of-speech tags in square brackets) by rank order after their index score.[23] The generic disparity in the number of marker words meeting the threshold for selection is immediately obvious: the histories use 144 words with greater degrees of consistency and exclusivity than the tragedies and comedies, for which only twenty-six and five marker words qualify respectively. Even so, the only word of dialogue in the corpus to be used consistently and exclusively by a single genre is *London*, receiving a perfect index of two as a history marker. This is

[22] J. F. Burrows, 'All the Way Through: Testing for Authorship in Different Frequency Strata', *Literary and Linguistic Computing*, 22 (2007), 27–47.

[23] See the documentation for MorphAdorner (morphadorner.northwestern.edu).

Table 21.2 Zeta marker words with an index score ≥1.5

	Marker words
Comedy (14) *v.* non-comedies (22)	1.623 *love*[n1-nn]; 1.578 *signior*[n1]; 1.513 *apparel*[n1], *error*[n1]; 1.500 *oaths*[n2].
Histories (10) *v.* non-histories (26)	2.000 *london*[n1-nn]; 1.962 *henry*[n1-nn]; 1.923 *richard*[n1-nn]; 1.900 *york*[n1-nn]; 1.885 *realm*[n1]; 1.846 *earl*[n1]; 1.823 *ashes*[n2], *march*[vvi]; 1.808 *john*[n1-nn]; 1.800 *harry*[n1-nn]; 1.762 *coronation*[n1]; 1.731 *english*[j-nn]; 1.723 *forces*[n2], *peers*[n2]; 1.708 *conquest*[n1]; 1.700 *bishop*[n1], *lancaster*[n1-nn], *plantagenet*[n1-nn], *salisbury*[n1-nn], *shrink*[vvi]; 1.685 *claim*[vvb], *george*[n1-nn], *reign*[n1]; 1.669 *allegiance*[n1], *government*[n1]; 1.662 *deposed*[vvn], *england's*[n1g-nn], *northumberland*[n1-nn]; 1.654 *england*[n1-nn], *france*[n1-nn], *fruit*[n1]; 1.646 *infant*[n1]; 1.631 *broken*[vvn], *colors*[n2], *highness'*[n1g], *kingly*[j]; 1.623 *arm*[vvi], *christendom*[n1-nn], *edward*[n1-nn], *gloucester*[n1-nn], *household*[n1], *ireland*[n1-nn], *mounted*[vvn], *thomas*[n1-nn]; 1.615 *king's*[n1g], *kings*[n2], *sound*[vvb], *subjects*[n2]; 1.608 *armed*[j_vn], *grievous*[j], *joined*[vvn], *league*[n1]; 1.600 *archbishop*[n1], *aspiring*[j_vg], *clarence*[n1-nn], *gaunt*[n1-nn], *jerusalem*[n1-nn], *levy*[vvi], *lineal*[j], *mayor*[n1], *warwick*[n1-nn]; 1.592 *foreign*[j], *zeal*[n1]; 1.585 *articles*[n2]; 1.577 *princely*[j]; 1.569 *castle*[n1], *chair*[n1], *guard*[vvb], *join*[vvb], *scepter*[n1], *shot*[n1]; 1.562 *chivalry*[n1], *conjunction*[n1], *glories*[n2], *gloucester's*[n1g-nn], *instead*[acp-p], *jesu*[n1-nn], *kent*[n1-nn], *kingdom's*[n1g], *peaceful*[j], *wales*[n1-nn]; 1.554 *french*[j-nn], *maintain*[vvi], *rose*[n1], *shoulders*[n2], *sovereign*[n1_j], *tempest*[n1]; 1.546 *bleeding*[j_vg], *parley*[n1], *spoil*[n1]; 1.538 *hollow*[j], *majesty*[n1]; 1.531 *cheerful*[j], *daily*[av_j]; 1.523 *blushing*[j_vg], *deputy*[n1], *dukedom*[n1], *hitherto*[av], *jades*[n2], *muse*[vvb], *pay*[n1], *proudly*[av_j], *signal*[n1], *submission*[n1], *victorious*[j], *william*[n1-nn]; 1.515 *loyal*[j]; 1.508 *dread*[j], *goods*[n2_j], *isle*[n1], *nobles*[n2_j], *prevail*[vvi], *prophesy*[vvi], *rebellion*[n1], *stoop*[vvb], *towns*[n2], *tread*[vvb], *treasons*[n2], *urged*[vvn]; 1.500 *albans*[n1-nn], *bolingbroke*[n1-nn], *brittany*[n1-nn], *canterbury*[n1-nn], *cardinal*[n1], *dauphin*[n1], *edward's*[n1g-nn], *exeter*[n1-nn], *giving*[vvg], *henry's*[n1g-nn], *insulting*[j_vg], *marched*[vvn], *norfolk*[n1-nn], *open*[vvb], *oxford*[n1-nn], *pomfret*[n1-nn], *prelate*[n1], *prophecies*[n2], *puissance*[n1], *rebels*[n2], *richard's*[n1g-nn], *saying*[vvg], *sovereign's*[n1g_j], *stafford*[n1-nn], *suffolk*[n1-nn].
Tragedies (12) *v.* non-tragedies (24)	1.667 *flies*[n2], *whore*[n1]; 1.583 *ope*[vvi], *safer*[j-c], *toad*[n1], *bidding*[n1_vg]; 1.542 *closet*[n1], *craves*[vvz], *drink*[n1], *strokes*[n2], *domestic*[j], *humanity*[n1], *slain*[vvn]; 1.500 *appetite*[n1], *capitol*[n1-nn], *cruel*[j], *drinks*[vvz], *general*[n1], *hail*[uh_v], *joint*[j], *nobly*[av_j], *senate*[n1], *served*[vvn], *silent*[j], *trouble*[vvb], *whereto*[crq-cs].

perhaps unsurprising, given that Shakespeare's histories are, as Ton Hoenselaars suggests, more properly 'English history plays', since the 'standard topographical gaze in the Folio category is English, with London as its ideological centre'.[24] Though they are set elsewhere, it is curious that the Folio comedies and tragedies never mention London in their dialogue—not even as a casual reference.

The history markers are predominately proper nouns, testifying to the genre's preoccupation with identifying multiple characters and shifting titles, though, as Table 21.2 shows, there are some interesting examples from other grammatical classes: the verbs *arm*, *claim*, *deposed*, *guard*, *join*, *levy*, and *shrink*, for instance, reflect the genre's thematic concern with asserting and maintaining political power. The comedy markers meeting the threshold index score are all nouns, most readily associated with the plays' frequent Italianate settings (*signior*) as well as the genre's primary motivator (*love*) and plot devices (*apparel*, *error*, *oaths*). More interesting are the marker words that are not always affixed to the generic category one might expect. While markers for tragedy reflect that genre's fixation on the baser aspects of human psychology (*appetite*, *craves*, *trouble*) as well as its grander scale (*humanity*), they also contain words that might be more reasonably associated with the military violence of the histories (*slain, strokes*). Likewise, while they may strike us as more germane to comedy or history, *whore* and *drink* are used more consistently in Shakespeare's tragedies, making them marker words for that genre.

The Zeta results suggest that Shakespeare's vocabulary is shared by all three Folio genres, with one exception: histories draw upon a relatively more distinctive vocabulary than comedies and tragedies but use only a single word (*London*) consistently and exclusively. Since Shakespeare's dramatic vocabulary is largely shared rather than exclusive, perhaps there are distinctive patterns in the way that plays of different genres use this common language. To answer this question, we count instances of the top 1,000 most frequent words across the corpus, calculated as a proportion of total tokens in each play (to enable valid comparison between texts of different lengths). For example, the determiner *the*—the most frequent word in the corpus—occurs 676 times in *Measure for Measure*, a play with 21,259 total tokens, so it has a proportion of 3.180. The result is a massive spreadsheet of thirty-six rows (one for each play), 1,000 columns (one for each of the top 1,000 most frequent words in the corpus), and the proportion counts for each word in each play.

Our data currently has 1,000 variables—far too many for human perception to visualize or easily comprehend—so we use Principal Component Analysis (PCA), a standard statistical procedure, to reduce data to its most salient factors. PCA works by transforming interrelated variables into new composite variables called 'principal components', which are uncorrelated with one another and 'ordered so that the first *few* retain most of the variation

[24] Ton Hoenselaars, 'Shakespeare's English History Plays', in *The New Cambridge Companion to Shakespeare*, eds. Margreta de Grazia and Stanley Wells (Cambridge: Cambridge University Press, 2010), 139.

present in *all* of the original variables'.[25] Unlike Zeta, where the algorithm needs to be 'supervised' by explicitly categorizing the data into reference and comparison sets (like First Folio genres and their opposites) before the process begins, PCA proceeds in an 'unsupervised' manner, transforming the data indiscriminately without any human intervention or pre-processing. Using PCA, we reduce our data to the first two principal components and, treating these scores as *x*- and *y*-coordinates respectively, project the plays as data-points into a two-dimensional scatterplot (Figure 21.3), later adding labels and symbols to identify the plays and their genres (since this information was withheld from the PCA algorithm and played no part in its calculations).

Reading the scatterplot in Figure 21.3, comedies are plotted as filled squares, histories as filled circles, and tragedies as crossed squares. The first principal component (PC1) on the *x*-axis represents the most important latent factor in the underlying relationships between the top 1,000 most frequent words in the plays, and the second principal component (PC2) on the *y*-axis the second most important (independent) latent factor. Within this two-dimensional space, the relative distances between plays reflect degrees of linguistic affinity: plays sharing similar proportions of the top 1,000 most frequent words cluster together, and vice versa. Thus, the greatest variance in the data is between *3 Henry*

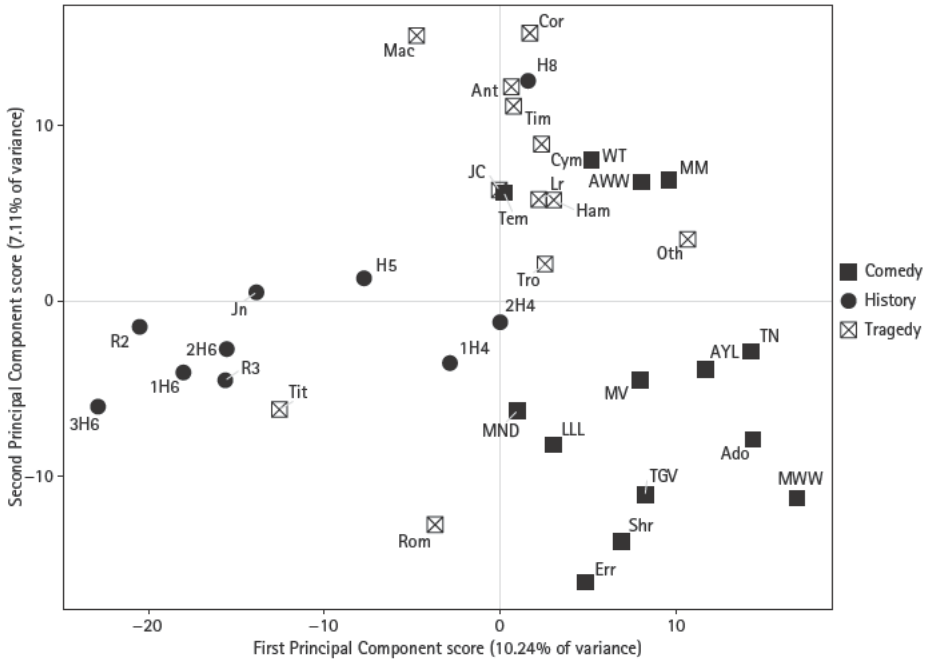


FIGURE 21.3 PCA scatterplot of First Folio plays using the top 1,000 most frequent words.

[25] I. T. Jolliffe, *Principal Component Analysis* (New York: Springer, 1986), ix; for a gentler, non-mathematical introduction, see Mick Alt, *Exploring Hyperspace* (Maidenhead: McGraw-Hill, 1990), 48–80.

VI and *Merry Wives* along PC1. A striking pattern emerges along the *x*-axis between histories, most of which score lower on PC1 and are plotted to the left, and comedies, which all score higher on PC1 and are plotted to the right. That Shakespeare's comedies and histories should present the biggest contrast in these terms may be counterintuitive, given the critical tendency to posit comedy and tragedy as polar opposites, but this result echoes the findings of previous quantitative studies.[26]

We may draw several generalizations from this PCA of Shakespeare's common dramatic vocabulary. Plays of the same Folio genre tend to cluster together, sharing similar proportions of the same words, though the divisions are imperfect and there are outliers. The most cohesive clusters are comedies and histories, which represent the greatest line of difference in dialogue between Folio plays. The outliers are interesting and warrant further investigation. *Titus Andronicus*, for example, was Shakespeare's first attempt at tragedy, a genre to which he would not return until much later; in the meantime, he turned his hand to writing English histories. Perhaps this explains why *Titus*, scoring the lowest on PC1 of all the tragedies, shows greater affinity with the histories and is plotted nearer them. Another anomaly is *Othello*, scoring the highest on PC1 of all the tragedies and plotted nearer the comedies. Does this reflect, as Jonathan Hope and Michael Witmore suggest, that *Othello* was built 'on structures that would ordinarily be employed in comedy' to heighten 'the emotional effect of down-turn' as 'things deteriorate' in the play?[27]

To explore these results in greater detail, we can inspect the correlations and 'weightings' of the underlying variables—that is, how individual words contributed to the principal components. As an example, we project personal pronouns into the same two-dimensional space to discern any meaningful patterns and correlations (Figure 21.4). Several patterns are observable. Older *th-* forms (*thee*, *thine*, *thou*, *thy*, *thyself*) are strongly correlated and weighted negatively on PC1, more characteristic of histories, where they are part of the formal, linguistically conservative parlance of the court when addressing individuals or groups of subordinate social rank. By contrast, modern forms of these pronouns (*you*, *your*, *yours*, *yourself*) are weighted positively on PC1, more characteristic of comedies. Histories also show a marked preference for some plural forms (*our*, *ours*, *us*, *we*), and comedies singular (*I, me, you, your, yours*). Except for *his*, all the gendered third person pronouns (*he, her*, *herself*, *him, himself, she*) are also weighted positively on PC1, reflecting the greater centrality of women in the comedies as both characters and plot devices. Similar patterns may be present in the relative use of other classes of words, but an exhaustive investigation is beyond the present scope.

[26] Hugh Craig, 'Stylistic Analysis and Authorship Studies', in *A Companion to Digital Humanities*, eds. Susan Schreibman, Ray Siemens, and John Unsworth (Malden: Blackwell, 2004), 273–288; Brett Greatley-Hirsch, 'Computational Studies', in *The Arden Research Handbook of Contemporary Shakespeare Criticism*, ed. Evelyn Gajowski (London: Arden Shakespeare, 2020), 205–221; Jonathan Hope and Michael Witmore, 'The Very Large Textual Object: A Prosthetic Reading of Shakespeare', *Early Modern Literary Studies*, 9 (2004), 6.1–36.

[27] Jonathan Hope and Michael Witmore, 'The Hundredth Psalm to the Tune of "Green Sleeves": Digital Approaches to Shakespeare's Language of Genre', *Shakespeare Quarterly*, 61 (2010), 374.

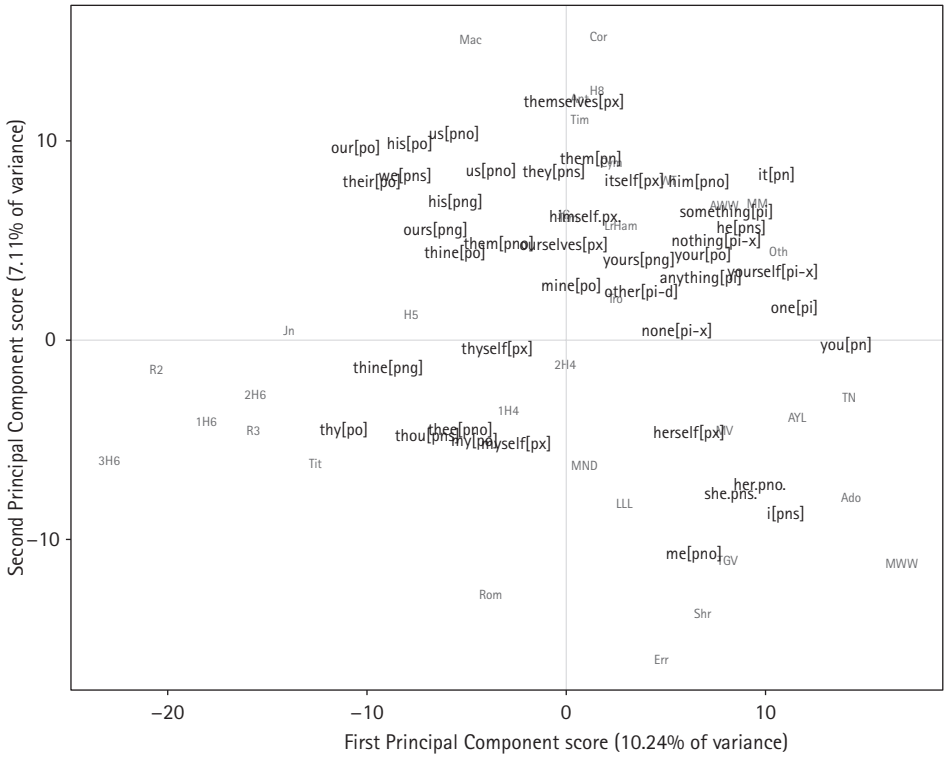


FIGURE 21.4 PCA biplot of First Folio plays using the top 1,000 most frequent words, selecting only personal pronouns.

In *Shakespeare's Dramatic Genres*, Lawrence Danson poses the following question: 'Do we know the genre [of a play] by its formal properties (what it looks like) or its affective properties (what it achieves in its audience)?'[28] To explore this question, and by way of a final experiment, we use sentiment analysis to plot the changing emotional valences in dramatic dialogue as it progresses through Shakespeare's Folio plays, sentence by sentence, and then compare the resulting plot 'shapes' within and across genres, thereby taking broad patterns in emotional language as a proxy for narrative structure.[29] At its most basic, sentiment analysis of this kind works by assigning pre-defined emotional values (e.g. from -1 for most negative to 1 for most positive, with 0 neutral) to words in a document, averaging the totals for a window of text (such as a sentence, a string of *n* words, a chapter, and so on), plotting these values in sequence and then 'smoothing' the trendline. It is an admittedly blunt heuristic tool, perhaps better suited to the marketing research for which it is routinely employed, made even more naïve because the available

[28] Lawrence Danson, *Shakespeare's Dramatic Genres* (Oxford: Oxford University Press, 2000), 3.

[29] See Jonathan Culpeper, Alison Findlay, Beth Cortese, and Mike Thelwall, 'Measuring Emotional Temperatures in Shakespeare's Drama', *English Text Construction*, 11 (2018), 10–37, and Katherine Elkins, *The Shapes of Stories: Sentiment Analysis for Narrative* (Cambridge: Cambridge University Press, 2022).

dictionaries defining the emotional valences of words do not accommodate historical languages. While the emotional values assigned may be entirely arbitrary and the algorithm may not account for the nuances of Early Modern English, sentiment analysis has the virtue of being systematic—that is, errors will be consistently and indiscriminately applied to all the plays.

With these caveats in mind, we used sentiment analysis to test whether plays identified in the First Folio as being the same genre adopt similar narrative structures (observed as broad patterns in emotional language) and produce graphs of similar shape. The algorithm we use is called sentimentr (github.com/trinker/sentimentr), an R package developed by Tyler Rinker to 'calculate text polarity sentiment' and which attempts to account for 'valence shifters' by statistically adjusting the values when certain words are used to intensify, lessen, or negate the emotional valence of a clause or adjacent term. Compare Demetrius's description of Hermia's rebuke, 'Lay breath so bitter on your bitter foe' (*MND* 3.5.44), which sentimentr assigns a negative score of -0.530, with his later drug-induced praise of Helena, 'O Helen, goddess, nymph, perfect, divine!' (3.5.138), which receives a strongly positive score of 0.714. For this experiment, we export the Folger Digital Texts to a plain-text format, stripping out everything except for words of dialogue and splitting the text of each play into grammatically complete sentences. The texts are ingested and processed by the sentimentr algorithm, which produces a smoothed trendline of emotional valence against narrative time (as a percentage of the total length) for every play in the order of printing in the First Folio (Figure 21.5).

Some plays share similar narrative 'shapes', but there is too much inconsistency for sentiment analysis to be a reliable discriminator of Shakespeare's comedies, histories, and tragedies. Not all tragedies, for example, follow the expected Aristotelian model of a great fall, starting positively and ending negatively: *Antony and Cleopatra* and *Romeo and Juliet* begin with a negative valence, whilst *Cymbeline*, *Julius Caesar*, and *King Lear* end positively. There are, however, interesting patterns that emerge within the variation when the 'shapes' of plays are observed in their 1623 printed sequence. The Folio publishers printed Shakespeare's histories in historical order: if we begin with the events of *Richard II* and end with *Richard III*, each play in succession picks up emotionally where the previous one left off, suggesting that Shakespeare paid explicit attention to the way these plays worked in sequence as he wrote and later revised these plays. That similar sentimental connections at the beginnings and endings are observed in series of comedies (*The Tempest* to *The Merry Wives of Windsor*, *The Comedy of Errors* to *Love's Labour's Lost*, *The Merchant of Venice* to *As You Like It*) and tragedies (*Titus Andronicus* to *Romeo and Juliet*, *Timon of Athens* to *Hamlet*, *Othello* to *Antony and Cleopatra*) raises the tantalizing possibility that the First Folio publishers ordered these genres with some consideration of their emotional effect.

Our stylometric experiments reveal that some distinctive features and patterns of language are common to the plays in each First Folio category, but they are also marked by high degrees of stylistic variation. We can broadly distinguish between plays of different genres by their language use, especially between comedies and histories, but anomalies, outliers, and overlaps show the Folio categories to have limited internal coherence and

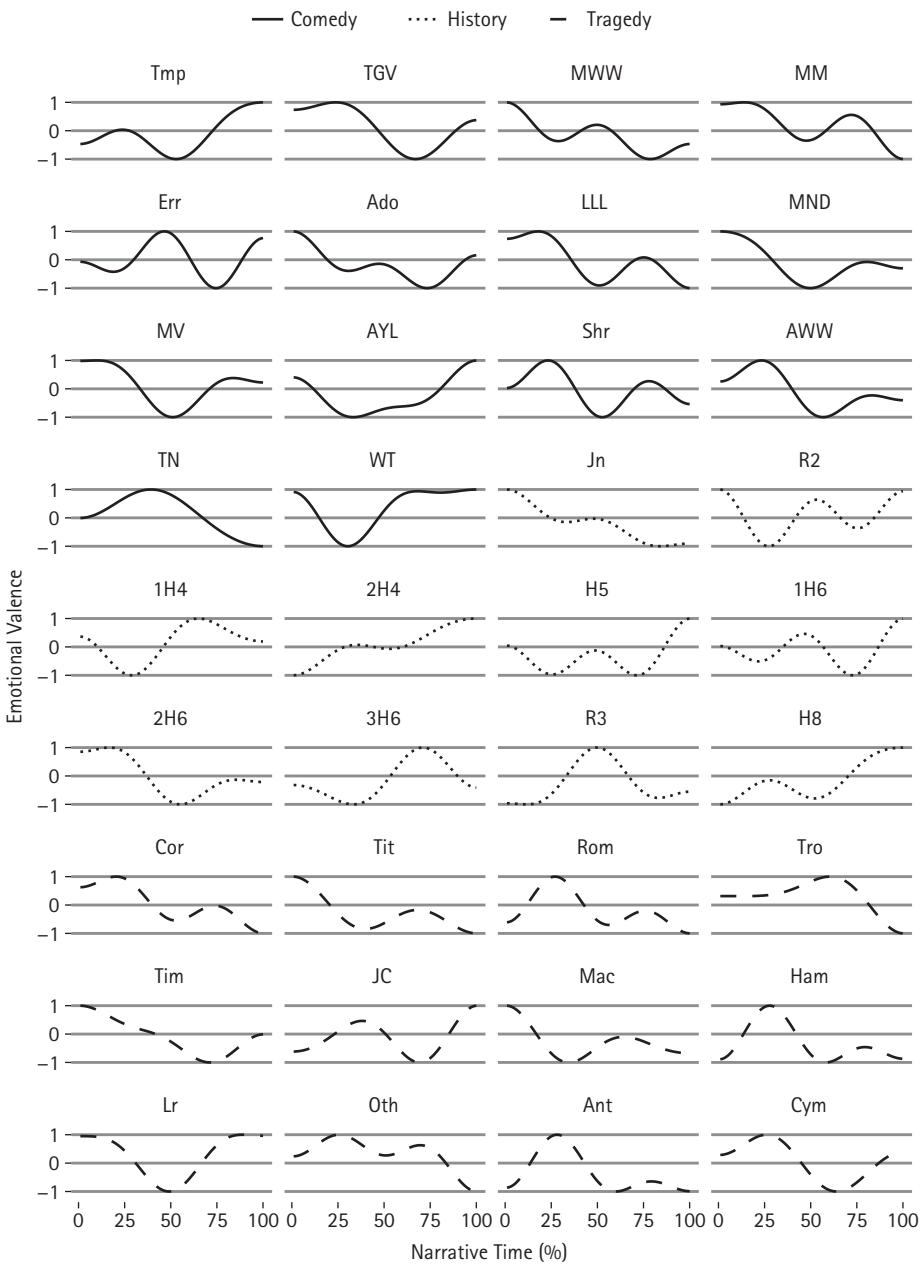


FIGURE 21.5 Sentiment analysis of First Folio plays progressing sentence by sentence.

cohesion. At what point do the labels of 'comedy', 'history', and 'tragedy'—generic skins acquired or discarded, as we have seen, in response to changing market demand—cease to be meaningful, useful, or relevant? If we read the corpus of Shakespeare's plays without the Folio's 'mark of genre', what other categories might emerge?

Shakespeare without Genre

In the previous section, we used computational methods to test systematically the dramatic language of First Folio genres for their distinctiveness, coherence, and cohesion. By way of conclusion, we shift the nature of the question from *classification* to *clustering*: rather than assessing how well plays fit predetermined genres, what can the groupings of Shakespeare's plays look like when free-form categories are allowed to emerge without recourse to external evidence? Are there coherent internal linguistic patterns that testify to Shakespeare's distinct habits of composition for each genre? For this final experiment, we rely on topic modelling algorithms to identify statistical relationships in language from which it is possible to derive any number of abstract *topics* with which to organize the data. (So, while the First Folio divides Shakespeare's plays into three fixed genres, topic modelling allows us to see how the plays might be allocated across any number of topics.) There are two crucial differences between genre categories and topics: whereas generic labels are predefined and imposed onto texts from outside (i.e. 'supervised'), topics are generated from within (i.e. 'unsupervised'); and if exclusivity is the organizing principle of genre (a text can be only one of comedy, history, or tragedy), the logic of topic modelling is the sort of inclusion that Polonius, with his 'tragical-comical-historical-pastoral', could easily grasp—every text is a mixture of topics, and every topic is a non-exclusive mixture of words.

To prepare the data, we relied on the Folger Digital Texts' tagging to replace the words of dialogue in the plays with their lemma or lexical root (thus, instances of *lose*, *loses*, *losest*, *loseth*, *losing*, and *lost* are all counted as *lost*) and generated a word-frequency table of all the lemmas across the corpus. As is routine in topic modelling, we then culled outliers so that our data only included lemmas common enough to be shared by multiple plays but also exclusive enough to be more characteristic of some plays than others. The topic modelling algorithm we use, Latent Dirichlet Allocation, assigns every lemma a probability score for its occurrence in each topic and then determines the proportion of topics to which each text belongs.[30] Mirroring the number of First Folio genres, we first instructed the algorithm to analyse our dataset of 3,367 lemmas to generate three topics. To avoid the tendency to assign generic bias, each topic is simply designated numerically. The resulting distribution was largely unsatisfying: as Figure 21.6 shows, each of the thirty-six Folio plays were distributed between all three topics with little variation. Three topics are insufficient for our model to produce a useful organizing principle.

In the hopes of locating a more significant result, and in imitation of the number of discrete generic labels initially mentioned by Polonius, we next instructed the algorithm to locate four topics. Figure 21.7 shows the proportion of each topic present in each play, and, given our explicit efforts to avoid generic bias, the results were surprising. It is

[30] David M. Blei, Andrew Y. Ng, and Michael I. Jordan, 'Latent Dirichlet Allocation', *Journal of Machine Learning Research*, 3 (2003), 993–1022.

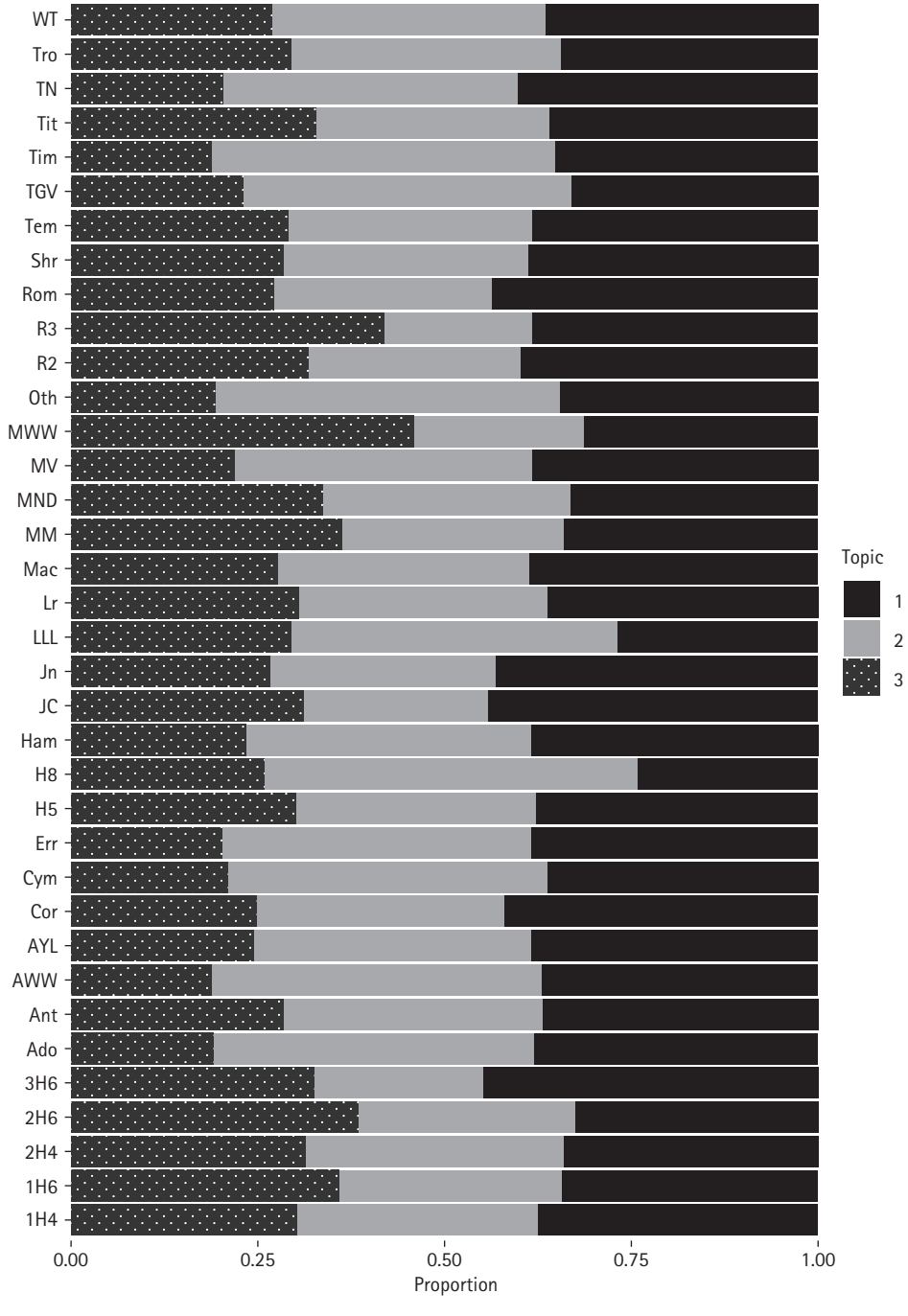


FIGURE 21.6 Allocation of First Folio plays across three topics.

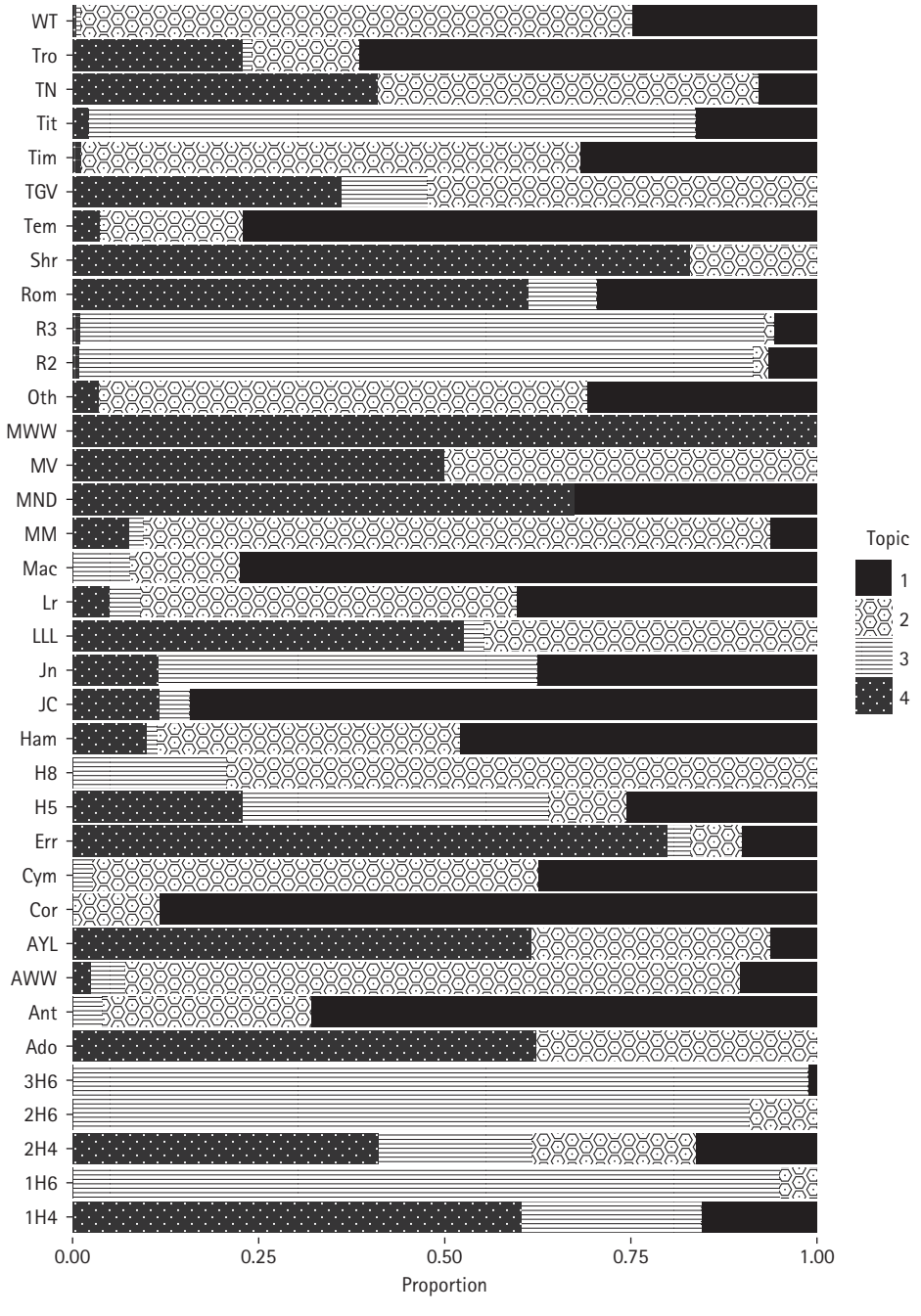


FIGURE 21.7 Allocation of First Folio plays across four topics.

important to remember that every play is a combination of all four topics, even if those proportions are too small to be represented visually at this scale, and that the distribution of topics is entirely unsupervised: the algorithm does not know to which generic category a play may belong. Even when evaluated exclusively through internal evidence, Shakespeare's plays do group together according to the Folio's established generic categories: by proportion, Topic 1 predominates plays designated as tragedies in the Folio, Topic 3 with histories, and Topic 4 with comedies. Yet this distribution was not evident when the algorithm was explicitly instructed to form three categories: it only emerged when the model was able to offer a fourth topic which, curiously, grouped together plays that critics have found most resistant to generic categorizations—the so-called 'problem plays', like *Measure for Measure* and *All's Well that Ends Well*. Examining the topic combinations also reveals that some plays are more similar to each other than we might otherwise assume (e.g. *The Winter's Tale* and *Timon of Athens*), while others are so largely associated with a single topic that they are stylistic outliers (e.g. *The Merry Wives of Windsor* and *3 Henry VI*). Other patterns are no doubt evident, and more granular research is needed to establish which are meaningful.

We used topic modelling to ask how the Folio might look if, rather than being externally imposed, the organization of the plays emerged from within according to shared patterns of language. The answer, as we have seen, is that few plays are predominantly associated with a single topic—most are sizeable combinations of multiple topics and some, like *2 Henry IV*, are especially 'balanced'. It may require some critical reorientation to make fuller use of topics as an organizing principle, but the adjustment may be 'worth the pains' (*MV* 2.6.33): rather than explaining away generic outliers like *Othello* or populating ever more granular micro-genres to accommodate the anomalous, topic models free us from externally imposed generic categories, and the logic of exclusion, revealing instead the plays' multiple points of connection.

Coda

Genre is foundational to the way critics, educators, readers, and audiences have approached Shakespeare's plays, but, as this chapter has shown, early modern dramatic genre was more complex and nuanced than the Folio's tripartite division of 'comedies', 'histories', and 'tragedies' suggests. The evidence of dramatic paratexts indicates that genre was flexible enough to serve the advertising needs of the print marketplace, while stylometric analysis both validates the Folio's generic boundaries and shows them to be porous. Computational methods allow us to explore Shakespeare's dramatic language at scale—to ask, 'What's genre to Shakespeare, or Shakespeare to genre?'—and, through techniques like topic modelling, promise new ways of seeing connections between the plays.

Bibliography

Alt, Mick, *Exploring Hyperspace* (Maidenhead: McGraw-Hill, 1990).

Arber, Edward, ed., *A Transcript of the Registers of the Company of Stationers of London; 1554–1640 A.D.,* 5 vols (London: Printed Privately, 1875).

Basu, Anupam, '"Ill shapen sounds, and false orthography": A Computational Approach to Early English Orthographical Variation', in *Early Modern Studies after the Digital Turn*, eds. Laura Estill, Diane K. Jakacki, and Michael Ulyot (Toronto: Iter Press, 2016), 167–200.

Bevington, David, ed., *The Complete Works of Shakespeare*, sixth edition (New York: Pearson, 2008).

Bourus, Terri, *Young Shakespeare's Young Hamlet: Print, Piracy, and Performance* (London: Palgrave, 2014).

Burrows, J. F., 'All the Way Through: Testing for Authorship in Different Frequency Strata', *Literary and Linguistic Computing*, 22 (2007), 27–47.

Craig, Hugh, 'Stylistic Analysis and Authorship Studies', in *A Companion to Digital Humanities*, eds., Susan Schreibman, Ray Siemens, and John Unsworth (Malden: Blackwell, 2004), 273–288.

Craig, Hugh and Brett Greatley-Hirsch, *Style, Computers, and Early Modern Drama: Beyond Authorship* (Cambridge: Cambridge University Press, 2017).

Culpeper, Jonathan, Alison Findlay, Beth Cortese, and Mike Thelwall, 'Measuring Emotional Temperatures in Shakespeare's Drama', *English Text Construction*, 11 (2018), 10–37.

Danson, Lawrence, *Shakespeare's Dramatic Genres* (Oxford: Oxford University Press, 2000).

Derrida, Jacques, 'The Law of Genre', *Critical Inquiry* 7 (1980), 55–81.

Dutton, Richard, *Shakespeare, Court Dramatist* (Oxford: Oxford University Press, 2016).

Edwards, Philip, ed., *Hamlet, Prince of Denmark,* updated edition (Cambridge: Cambridge University Press, 2003).

Elkins, Katherine, *The Shapes of Stories: Sentiment Analysis for Narrative* (Cambridge: Cambridge University Press, 2022).

Farmer, Alan B., 'Shakespeare as Leading Playwright in Print, 1598–1608/09', in *Shakespeare and Textual Studies*, eds. Margaret Jane Kidnie and Sonia Massai (Cambridge: Cambridge University Press, 2015), 87–104.

Farmer, Alan B. and Zachary Lesser, '"Vile Arts": The Marketing of English Printed Drama, 1512–1660', *Research Opportunities in Renaissance Drama*, 39 (2000), 77–165.

Farmer, Alan B. and Zachary Lesser, ed., *DEEP: Database of Early English Playbooks*: deep.sas.upenn.edu.

Greatley-Hirsch, Brett, 'Computational Studies', in *The Arden Research Handbook of Contemporary Shakespeare Criticism*, ed. Evelyn Gajowski (London: Arden Shakespeare, 2020), 205–221.

Higgins, Ben, *Shakespeare's Syndicate: The First Folio, its Publishers, and the Early Modern Book Trade* (Oxford: Oxford University Press, 2022).

Hoenselaars, Ton, 'Shakespeare's English History Plays', in *The New Cambridge Companion to Shakespeare*, eds. Margreta de Grazia and Stanley Wells (Cambridge: Cambridge University Press, 2010), 137–151.

Hope, Jonathan and Michael Witmore, 'The Very Large Textual Object: A Prosthetic Reading of Shakespeare', *Early Modern Literary Studies*, 9 (2004), 6.1–36.

Hope, Jonathan and Michael Witmore, 'The Hundredth Psalm to the Tune of "Green Sleeves": Digital Approaches to Shakespeare's Language of Genre', *Shakespeare Quarterly*, 61 (2010), 357–390.

Jolliffe, I. T., *Principal Component Analysis* (New York: Springer, 1986).

Kesson, Andy, 'Was Comedy a Genre in English Early Modern Drama?', *British Journal of Aesthetics*, 54 (2014), 213–225.

Kewes, Paulina, 'The Elizabethan History Play: A True Genre?', in *A Companion to Shakespeare's Works*, Vol. 2, eds Richard Dutton and Jean E. Howard (Malden: Blackwell, 2003), 170–193.

Shannon, C. E., 'A Mathematical Theory of Communication', *Bell System Technical Journal*, 27 (1948), 379–423.

Shannon, C. E., 'Prediction and Entropy of Printed English', *Bell System Technical Journal*, 30 (1951), 50–64.

Smith, Ian, *Black Shakespeare: Reading and Misreading Race* (Cambridge: Cambridge University Press, 2022).