

The Misuse of Function Words in Shakespeare Authorship Studies

Authorship attribution studies aim to identify the author of an anonymously published or collaborative work by identifying one or more linguistic features that discriminate them from other writers. Since their beginning in the early nineteenth century attribution scholars have studied many different aspects of drama, poetry and prose. Some approaches have proved unreliable, such as those studying sentence-lengths; the varying percentages of nouns, verbs and other parts of speech; the position of word-occurrences within a sentence; and the ratio between word-types (lexical units) and word-tokens (the number of occurrences).¹ Other approaches have proved useful, such as those studying verse structure: feminine endings (decasyllabic lines having an extra syllable), the position of breaks or pauses within a verse line; and the proportion of run-on lines.² Other valuable methods include authorial linguistic preferences such as the use of word-contractions and expletives,³ and the tendency of authors to repeat favourite phrases or collocations. This approach has been dismissed by exponents of other methods as ‘hunting for parallel passages’,⁴ but it has been given a new lease of life by the availability of large electronic corpora (such as LION and EEBO), and by the application of anti-plagiarism software.⁵ These are only a few of the competing systems currently in use. Indeed, for the two well-informed scholars who have periodically reviewed authorship attribution studies, Joseph Rudman and Patrick Juola, the scene is characterized as a chaos of conflicting systems, with a damaging lack of agreement on many fundamental issues.⁶ Other

¹ For a useful survey of those methods see David I. Holmes, ‘Authorship attribution’, *Computers and the Humanities*, 28 (1994): 87–106.

² See Brian Vickers, *Shakespeare, Co-Author. A Historical Study of Five Collaborative Plays* (Oxford, 2002), 47–57.

³ *Ibid.*, 80–90.

⁴ *Ibid.*, 57–75.

⁵ See Brian Vickers, ‘Identifying Shakespeare’s additions to *The Spanish Tragedy* (1602): a new(er) approach’, *Shakespeare* 8 (2012): 13–43, and ‘Shakespeare and the 1602 Additions to *The Spanish Tragedy*: a method vindicated’, *id.* 12 (2016); ‘The two authors of *Edward III*’, *Shakespeare Survey* 67 (2014): 69–84; ‘Giving Greene his *Groatsworth*’, *Papers of the Bibliographical Society of America* (forthcoming).

⁶ See, e.g., Joseph Rudman, ‘The state of authorship attribution studies: some problems and solutions’, *Computers and the Humanities*, 31 (1988): 351–65, and ‘The state of non-traditional authorship attribution

scholars are more sanguine about recent developments, heralding ‘a new era of authorship analysis technology’.⁷

The first rigorously statistical method in modern attribution studies was Alvar Ellegård’s work on the authorship of the *Junius Letters*, a series of political pamphlets written in 1769–72.⁸ Ellegård investigated the candidate identified by John Taylor in 1813, Sir Philip Francis, a Whig politician whose career overlapped with many of the events treated in the pamphlets. The *Junius* corpus ran to over 200,000 words, slightly less than Francis’s, and Ellegård described the method he used as ‘mainly a vocabulary test’. He defined two groups of expressions (including ‘both individual words, phrases, and even grammatical constructions’): ‘plus words’, ‘used much more often by the writer in question than his colleagues’, and ‘minus words’, those which he used much less often (1962:12). In order to construct a testing list, Ellegård had to establish both the relative frequencies of the target author’s vocabulary, but also ‘the average relative frequencies of the vocabulary appearing in literature of the same general kind, and belonging to the same period as our author’ (20). So he created a preliminary testing list by handpicking the words that ‘seemed likely to have an effective distinctiveness ratio’ (21–2). He ‘carefully combed the Junius material for words and expressions’ which seemed to occur with remarkable frequency, and ‘did the same for the comparative sample of a million words, taken from about a hundred different authors’, noting both frequent and infrequent occurrences. This gave him ‘a preliminary list of Junian plus words’ and ‘Junian minus words’, amounting to 458 expressions, which he then

studies – 2012: some problems and solutions’, *English Studies*, 93 (2012): 259–74. In the latter study Rudman noted ‘a muddled but gradual advancement in some areas’, but at the same time ‘a slew of new, unproven methodologies, new problems, and new controversies have the field in turmoil’ (p. 259). Patrick Juola, in his magisterial survey, ‘Authorship attribution’, *Foundations and Trends in Information Retrieval* 1 (2006): 234–334, similarly lamented that ‘the state of the art is confusing’ (233), and ‘the field itself is somewhat in disarray, with little overall sense of best practices and techniques’ (235), there being ‘an almost limitless number of techniques’ (246).

⁷ Efstathios Stamatatos, ‘A survey of modern authorship attribution methods’, *Journal of the American Society for Information Science and Technology*, 60 (2009): 538–56 (539).

⁸ See Alvar Ellegård, *A Statistical Method for Determining Authorship: the Junius Letters, 1769–1772* (Göteborg, 1962), also *Who was Junius?* (Stockholm, 1962), a non-technical monograph discussing the historical and literary aspects of the Junius problem.

memorized, having read each page of text twice (22–3; Appendix II). These manual operations needed extraordinary diligence, to which he added statistical expertise (Appendix III), and finally used a computer to produce a mass of data (Appendix IV), confirming Francis as the author. This hybrid method, admirable in its thoroughness, has never been used again. Ellegård’s successors have avoided reading texts, and relied on computers from the outset.

The most popular method has been the study of authors’ varying use of function words – that is, those parts of speech which have a grammatical function, such as pronouns, prepositions, conjunctions, articles. This method came to prominence through its use in the study by Frederick Mosteller and David Wallace of the authorship of *The Federalist Papers*.⁹ Its principal exponent in recent years has been John Burrows, who published a study of Jane Austen’s novels in 1987, followed by a series of essays.¹⁰ On some occasions he has been joined by his associate and former pupil, Hugh Craig.¹¹ Another notable follower of Burrows is David Hoover, who has applied his methods to nineteenth-century novels.¹² Despite the general approval within the scholarly community of using function words for authorship attribution, in this paper I shall argue that a central plank of its theoretical justification is mistaken, which makes it unsuitable for the study of Shakespeare and other early modern dramatists.

⁹ See Frederick Mosteller and David Wallace, *Inference and Disputed Authorship. ‘The Federalist’* (Reading, MA, 1964). I quote from the facsimile reprint (Stanford, CA, 2001), with an introductory essay by John Nerbonne, ‘The exact analysis of text’ (pp. xi–xx).

¹⁰ See John Burrows, *Computation into Criticism: A Study of Jane Austen’s Novels and an Experiment in Method* (Oxford, 1987); ‘“An Ocean Where Each Kind...”: statistical analysis and some major determinants of literary style’, *Computers and the Humanities*, 23 (1989): 309–21; ‘“Delta”: a measure of stylistic difference and a guide to likely authorship’, *Literary and Linguistic Computing*, 17 (2002): 267–87; ‘Questions of authorship attribution and beyond’, *Computers and the Humanities*, 37 (2003): 3–32; ‘All the way through: testing for authorship in different frequency strata’, *Literary and Linguistic Computing*, 22 (2007): 27–47.

¹¹ See, e.g., Hugh Craig and Arthur F. Kinney, *Shakespeare, Computers, and the Mystery of Authorship* (Cambridge, 2009), hereafter referred to as ‘*Mystery of Authorship*’; Hugh Craig and John Burrows, ‘A collaboration about a collaboration: the authorship of *King Henry VI, Part Three*’, in Marilyn Deegan and Willard McCarty (eds.), *Collaborative Research in the Digital Humanities* (Farnham and Burlington VT, 2012), pp. 28–65.

¹² See D. L. Hoover, ‘Statistical stylistics and authorship attribution: an empirical investigation’, *LLC*, 16 (2001): 421–44; ‘Frequent word sequences and statistical stylistics’, *id.*, 17 (2002): 157–80; ‘Frequent collocations and authorial style’, *id.*, 18 (2003): 261–86; ‘Testing Burrows’s Delta’, *id.*, 19 (2004): 453–75.

I

Mosteller and Wallace studied a series of short essays published in *The Federalist* in 1787–8 signed ‘Publius’, which attempted ‘to persuade the citizens of the State of New York to ratify the Constitution’ (1964: 2). Three authors contributed these papers, Alexander Hamilton, John Jay and James Madison. The authorship of most papers is known, apart from twelve where the attribution attempts resulted in a straight contest between Hamilton and Madison. Initially, Mosteller and Wallace hoped to distinguish them according to their choice of words – Hamilton used *while* where Madison in a corresponding situation used *whilst*, and to the frequency with which they used certain words – Hamilton used both *upon* and *enough* more frequently than Madison (11). Mosteller and Wallace found that they could not distinguish the two authors by their use of contextual words, those defined by the subject matter of the essays, for each used many of the same terms. They therefore decided to study high-frequency words, especially the class known as ‘*function* words’, which they defined as ‘the filler words of the language’, comprising ‘prepositions, conjunctions, pronouns, and certain adverbs, adjectives, and auxiliary verbs’ (17). Mosteller and Wallace compiled a list of fifty-one frequently occurring function words in the disputed papers (28–30), and studied their rate of occurrence in an exhaustive series of tests. This model statistical study ultimately ascribed all twelve papers to Madison. Their whole enterprise aimed to distinguish the styles of two authors writing persuasive prose, and for this problem they discovered that function words ‘are a fertile source of discriminators’, while ‘context is a source of risk. We need variables that depend on authors and nothing else’ (265). But while succeeding in their aim Mosteller and Wallace did not discuss why Hamilton used *upon* and *enough* so frequently. Was it a conscious or unconscious choice?

Subsequent attribution scholars came to regard the use of function words as revealing an author's unconscious preferences. David Holmes endorsed R.W. Bailey's recommendation that the writings chosen for comparison should be of a sufficient length 'to reflect the linguistic habits of the author of the disputed text and also those of each of the candidates'. The comparison should concentrate on 'quantifiable features of the texts', having these 'general properties...: "They should be salient, structural, frequent and easily quantifiable and relatively immune from conscious control"' (1994: 88). The phrase 'relatively immune from conscious control' led to this becoming a key criterion for later studies. Holmes added a further requirement: 'For discrimination purposes we need context-free or "function" words' (90). David Hoover praised the work of John Burrows on the early nineteenth-century novel, relying on 'the frequencies of words such as *the, and, of, to*' as differentiating 'authors, novels, and even characters within a single novel from each other'. Hoover commented that 'Precisely because such words are so frequent in English and so *unlikely to be regulated by authors*, their frequencies may reflect *authorial habits* that remain constant in spite of differences in subject matter, point of view, or theme' (2002, 157). In a later essay Hoover proposed to 'analyse only collocations of function words, *avoiding words that might be consciously manipulated by an author and concentrating on those that are more likely to be habitual...*' (2003, 266).¹³ I have italicized both instances of this emerging theory that authors are unconscious of their choices of function words, thus making them reliable linguistic markers. John Nerbonne has recently summarized development in attribution studies since Mosteller and Wallace, singling out Burrows as having helped to bring about the 'remarkably widespread acceptance that high-frequency elements provide the most reliable clues, reflecting, as they do, unconscious tendencies of authors'. As he commented,

¹³ Cf. also Hoover's description of the 'traditional view' that 'the most frequent words, typically function words, are likely to be beyond the author's control and are therefore suitable for authorship attribution' (2004, 470).

‘it is striking that Mosteller and Wallace’s conclusion still stands, but even more striking that exactly the same reason is offered after so many years...’ (2007: A–xvii).

Unfortunately, this ‘reasoning’ is doubly unsatisfactory. First, it is impossible to verify, given that authors’ ‘unconscious tendencies’ cannot be examined. Secondly, it ignores the fundamental nature of language. I suggest a different explanation, namely that the choice of function words is determined by a larger set of conscious choices, the language user’s intended communication of meaning. Prepositions, for example, fulfil the function of connecting meaning-bearing words according to the norms of grammatical usage, so that the speaker’s meaningful utterance is correctly communicated. A standard grammar of contemporary English, prepared by scholars connected with the pioneering survey of English usage set up by Randolph Quirk at University College London, defines ‘Prepositional meanings’ in these terms:

In the most general terms, a preposition expresses a relation between two entities, one being that represented by the prepositional complement. Of the various types of relational meaning, those of PLACE and TIME are the most prominent and easy to identify. Other relationships such as INSTRUMENT and CAUSE may also be recognized although it is difficult to describe prepositional meanings systematically in terms of such labels. ... So varied are prepositional meanings that no more than a presentation of the most notable semantic similarities and contrasts can be attempted here.¹⁴

The phrase ‘prepositional meanings’ instantly reveals the dimension ignored in stylometric computations, the role that prepositions fulfil in the broader communication of meaning. The authors of this *Grammar* show the ‘relations of meaning between a number of prominent prepositions of place’ can be clarified by distinguishing between ‘destination’ and ‘position’ in three dimension types: point, line or surface, and area or volume. The positive prepositions in the first category are *to* (destination) and *at* (position), their negatives *away* and *away from*. In the second category, line or surface, the positive terms are *on* and *on (to)*, the

¹⁴ Randolph Quirk, Sidney Greenbaum, Geoffrey Leech, Jan Svartvik, *A Grammar of Contemporary English* (London, 1972), p. 305. Future page references are incorporated into the text.

negatives for both *off*, while in the third (area or volume) they are *in(to)* and *out of*. As they define it, ‘the dimension-type of a preposition is the dimensional property ascribed, subjectively speaking, to the location *denoted* by the prepositional complement’ (308; my italics).

The fundamental property of prepositions is to denote aspects of space, time and causation from the subjective viewpoint of the speaker, whether the writer of a sole-authored composition, or one of the many characters in a play. ‘Apart from simple position, prepositions may express RELATIVE POSITION of two objects or groups of objects’ (310) – always from the speaker’s viewpoint. Different prepositions denote differences within these relative positions: ‘*over* and *under* tend to indicate a direct vertical relationship and/or spatial proximity, while *above* and *below* may indicate simply “on a higher/lower level than”’ (313). The choice between these pronouns may or may not be a conscious one for the speaker, but it is made in accordance with the utterance’s intended meaning. A word used frequently in this section of the *Grammar* is the ‘idea’ that the speaker wishes to communicate: ‘With verbs of motion, prepositions may express the idea of PASSAGE (i.e. movement towards and then away from a place) as well as destination’, by using such prepositions as *over*, *under*, *behind* and *past*. Another group of pronouns, including *up*, *down*, *along*, *across*, are used ‘with verbs of motion ... expressing movement with reference to a directional path’ (313). Most pronouns of this type imply ‘a “point of orientation”, at which (in reality or imagination) the speaker is standing. *Beyond* (= “on the far side of”) is a preposition whose primary meaning is one of orientation’ (314).

The second type of pronouns discussed in this *Grammar of Contemporary English* are those relating to time. They have ‘only two “dimension-types”, viz. “point of time” (*at*, *on*) and “period of time”’ (*for*, *over*, *from ... to*) (317), with correspondingly limited areas of meaning. The third type, however, concerning prepositional phrases used as adjuncts (that is,

integrated within the clause), offers ‘a range or spectrum of meanings’. The *Grammar* distinguishes two ‘fields of prepositional meaning’. The first consists of ‘cause–purpose’ prepositions, ‘expressing either the material cause’, as in ‘*Because* of the drought...’ or ‘the psychological cause (motive)’, as in ‘*out of* gratitude’ (321). If a speaker wishes to describe a purpose, or intended destination, she can use *for*, as in ‘He’ll do anything *for* money’, or ‘everyone ran *for* safety’ (321). When *for* is followed by a noun phrase denoting persons, ‘the meaning is rather one of “intended recipient”’, as in ‘He laid a trap *for* his enemies’ (322). The second ‘field of meaning’ consists of ‘means–agentive’ prepositions, which cover aspects of manner by using *with* (‘We were received *with* the utmost courtesy’), or *like* (‘*Like* a pestilence, the army swept through the city...’). But ‘*like*’ can also be used purely to convey resemblance, as in ‘*Life is like* a dream’ (319). A writer may use ‘*like*’ subconsciously, but the intended meaning will always govern the choice. The preposition *with*, ‘especially when followed by an animate (human) complement ... has the meaning “in company with” or “together with”’, the negative being ‘without’. Three common prepositions communicate the choice between support or opposition: “*For* conveys the idea of support; *with* that of solidarity or movement in sympathy; *against* conveys the contrary idea of opposition. In this use, there is no negative *without* contrasting with *with*’ (326). In all these cases the choice of preposition, whether or not freely available to the speaker’s conscious or subconscious mind, is subordinate to the idea or intentional meaning being communicated.

Some practitioners of function-word counting show an awareness of a wider linguistic horizon. For John Burrows, studying ‘embedded histories’ in eighteenth-century novels, ‘Henry Fielding’s predilection for *which* points to a different kind of syntax from that signified by Sarah Fielding’s predilection for *but* ... and his use of *the*, *a*, and *an* points to a greater emphasis than hers on *things*’. (1989: 320). In a later essay Burrows offered ‘to

arrange the commonest words of the language in loose grammatical categories', proposing four classes:

1. *Referential*, including articles and personal pronouns
2. *Temporal/modal*, including auxiliary verbs and appropriate adverbs
3. *Connective*, including conjunctions, prepositions, and relative pronouns
4. *Modificatory* including adjectives and adverbs

(2003: 28)

Of the four, Burrows described the *modificatory class* as 'the least interesting', although he noted that Aphra Behn frequently used 'the full range of common absolutes like *every, no, too, and all*', while Henry Fielding 'liked *very*'. Following Mosteller and Wallace, Burrows found that 'the *referential class* is made volatile by the presence of the personal pronouns',¹⁵ and judged that 'the inflected auxiliary verbs bring the same sort of volatility to the *temporal/modal class*. These verbs and the temporal adverbs shed particular light on texts where the present is either embraced or else avoided in favour of reminiscence or desire' (29). Having eliminated the less valuable categories, Burrows kept the most important till last:

Mostly by virtue of the stability and unexpected individuality that attach to the use of prepositions, the *connective class* is much the strongest of the four in identifying authorship. As a group, the prepositions are strong markers of noun-laden, 'thingish' texts. But, within the group, there are contrasts between a propensity to emphasize abstract ideas in idioms founded upon *of, by, and from* and a propensity to emphasize activity in idioms founded upon *on, upon, and into*. (29)

These were no more than suggestive observations, but they showed that Burrows was aware of a larger linguistic domain, linking function words both to the material world and to 'abstract ideas'.

¹⁵ Mosteller and Wallace 'eliminated ... personal pronouns except in "the possessive form"', observing that 'certain types of function words are potentially dangerous. Personal pronouns and auxiliary verbs, especially with respect to mood and tense, are likely to be related to external details, and inference from them is difficult' (39).

Hugh Craig's justifications for basing his analyses on function-word frequencies are narrower. His procedure is based on 'word-counting',¹⁶ words being 'the lowest level of sense unit in language' (11). Craig follows Mosteller and Wallace, the model for Burrows,¹⁷ in choosing words 'independent of local subject matter', notably 'function words ... [which] have vital grammatical functions but little or no semantic content' (13, 26). He compiled a list of 200 function words (29, 221–2). Craig is evidently not interested in semantics, and in his account of language never mentions the notion of meaning (1–8). He refers to 'communication' (8) but not to what is communicated, or by whom. Language is denuded of its reason for being. As Roman Jakobson often observed, 'The study of language without meaning is meaningless'.

THE RELEVANCE OF GENRE

It is a commonplace in attribution studies that a given method does not necessarily work on any text. David Holmes observed that

All authorship studies begin with a choice of criteria believed to characterize authors. One should probably not believe that any single set of variables is guaranteed to work for every problem, so researchers must be familiar with variables that have worked in previous studies as well as the statistical methods to determine their effectiveness for the current problem. (1994: 104)

Joseph Rudman's recommendations for an 'Experimental Plan' for attribution studies begins: 'When formulating the experimental design, the practitioners must make sure they are testing for authorship and not some other variable such as genre...' (2012: 266). Patrick Juola has commented that 'the inherent accuracy' of authorship attribution techniques is subject to several critical issues, notably 'genre, representativeness and corpus size'. A large authorial corpus, extending to 'millions of words', offers great scope, but 'the document of interest

¹⁶ Craig, *Mystery of Authorship*, (xvii, xviii).

¹⁷ In the Preface Craig states that 'computational stylistics stems from the work of John Burrows, beginning in the 1980s' (xv), forgetting that Mosteller and Wallace showed him the way.

may be in a completely different and incompatible genre. Given the well-understood differences in the statistics, and therefore fingerprints, across different genres,¹⁸ identifying an accurate *cross-genre* technique is an even more difficult task' (Juola 2006: 246; author's emphasis). Or, as he succinctly put it, research findings 'may be more applicable to some forms of texts than others' (286).

As we have seen, the two foundational studies in modern statistical attribution studies were based on single-author texts. Ellegård worked on the *Letters* of Junius, a series of political papers by one author, whom he identified as Philip Francis. Mosteller and Wallace's study of the *Federalist* papers focused primarily on two of the three authors, Alexander Hamilton and James Madison. Although in this case two authors were involved, each of them, like Francis, wrote expository prose from a single perspective, a single point of view using an individual linguistic system, or idiolect. At first sight John Burrows was a welcome exception to this single-author focus. In the research programme leading up to his 'Delta' method he created 'a database by twenty-five poets of the English Restoration period', amounting to some 200 poems of varying lengths, from 174 to 6812 words (2002: 269, 284–7). The poems chosen embrace a wide range of genres: dramatic prologues and epilogues (Congreve, Prior); elegies and epitaphs (Cowley, Oldham, Gould, Phillips, Waller); translations from the French (Cotton, Phillips); paraphrases and imitations of Ovid and Horace (Congreve, Cowley, Oldham, Prior); Pindaric odes (Cowley); familiar epistles (Congreve, Cotton, Cowley, Gould, Phillips, Swift, Waller); panegyrics (Cowley, Waller); religious poems (Gould, Phillips); love poems (Prior); literary criticism (Swift, Waller); political poems (Cowley, Swift, Waller); and the largest category, satires (Cotton, Gould, Oldham, Swift).

¹⁸ Juola cited D. Biber, S. Conrad and R. Reppen, *Corpus Linguistics: Investigating Language Structure and Use* (Cambridge, 1998).

There is a pleasing diversity of tones and poetic forms in this corpus. Yet they share, with the *Junius Letters* and the *Federalist* papers, one important feature: they are all spoken by one voice (whether that of the author or of a *persona*) and represent a single point of view. Burrows selected the 150 most common words in the corpus (2002: 274), and performed meticulous computations, claiming that ‘the Delta procedure’ provides ‘a direct guide to likely authorship’ (276). I think that Burrows claims too much when he writes that his method can highlight ‘the differences between one style and another’ (271), and establish ‘stylistic signatures’ (277), since the restricted lexicon he uses covers only a tiny aspect of what we mean by style. Yet it evidently succeeds at the micro-analytical level, given a database limited to poems having a single viewpoint.

In drama the situation is totally different. Apart from liminary matter, such as dedications, discursive prefaces, prologues and epilogues (when securely attributed), a dramatist does not write *in propria persona*. Ben Jonson, of course, is a partial exception to this rule, popping up in prefatory discussions, identifying himself with some characters, and even (according to reports) being present at performances, grimacing when the actors did not perform according to his wishes. But he carefully individualised his characters: the language of Epicure Mammon is far removed from that of Tribulation or Kastil.¹⁹ For Shakespeare we have only two documents written *in propria persona* (excluding his will, which follows legal conventions), the dedications to his narrative poems, both addressed to Henry Wriothesley, Earl of Southampton. Interestingly enough, given our concern with authorial styles, both display the type of rhetorically structured prose that recurs throughout his plays. That to *Venus and Adonis* (1593) is more complex:

*TO THE RIGHT HONORABLE HENRY WRIOTHESLEY,
EARL OF SOUTHAMPTON, AND BARON OF TICHFIELD.*

¹⁹ For an admirable account of Jonson’s methods of linguistic differentiation see Arthur H. King, *The Language of Satirized Characters in Poëtaſter, a socio-stylistic analysis 1597-1602* (Lund, 1941).

RIGHT HONORABLE,

I KNOW not how I shall offend in dedicating my unpolished lines to your lordship,
nor how the world will censure me for choosing

so strong a prop to support
so weak a burden;

only, if your honour seem but pleased,

I account myself highly praised,
and vow to take advantage of all idle hours,
till I have honoured you with some graver labour.

But if the first heir of my invention prove deformed,

I shall be sorry it had so noble a god-father,
and never after ear so barren a land,
for fear it yield me still so bad a harvest.

I leave it to your honourable survey,

and your honour to your heart's content;
which I wish may always answer your own wish
and the world's hopeful expectation.

Your honour's in all duty,

WILLIAM SHAKESPEARE.

The prose style of this courteous dedication deploys various types of symmetry. The syntax uses parallel structures: 'I know not how ... nor how'; pivoting phrases that can extend to sequences of four words, while expressing a single antithesis ('so strong a prop' / 'so weak a burden') or a double one ('so noble a god-father' / 'so barren a land' / 'so bad a harvest'). Shakespeare places significant matching phrases at the end of members, single and double: 'pleased' / 'praised; 'idle hours' / 'graver labour'. He varies the form of words ('honourable' / 'honour'), and expands the parallel phrases to form a satisfying conclusion: 'your heart's content' / 'your own wish' / 'the world's hopeful expectation'.

His dedication of *Lucrece* (1594) is more concise, and more highly patterned:

*TO THE RIGHT HONORABLE HENRY WRIOTHESLY,
Earl of Southampton, and Baron of Tichfield.*

The love I dedicate to your lordship is without end; Whereof
this pamphlet, without beginning, is but
a superfluous moiety.

The warrant I have of your honourable disposition,
not the worth of my untutored lines,
 makes it assured of acceptance.

What I have done is yours;
what I have to do is yours; being part in all
 I have, devoted yours.

Were my worth greater,
 my duty would show greater;
meantime, as it is,

 it is bound to your lordship,
 to whom I wish long life,
 still lengthened with all happiness.

Your lordship's in all duty,

WILLIAM SHAKESPEARE.

Every word in that sequence is part of a carefully structured phrase or clause, eloquent and courteous. The style is appropriate to the genre, courtly acknowledgement of a patron. The same rhetorical figures of balance and opposition – *parison*, *anaphora*, *epistrophe*, *polyptoton*, *anadiplosis* – recur throughout the prose in Shakespeare's plays.²⁰ But there they become devices used by a host of characters, each for their own personal ends, and are supra-authorial, so to speak. Like prepositions, conjunctions, adverbs and other common words, the rhetorical figures fulfil their function in a meaningful discourse. Extracting them, or function words, from that huge variety of contexts in order to claim them as authorial style markers would be a reductive act that annuls their function.

²⁰ See, e.g., Brian Vickers, *The Artistry of Shakespeare's Prose* (London, 1968; second edition, London, 2005).

The crucial point that we must keep in focus is that the words in Shakespeare's plays are not his but those of his characters. They were written by him, of course, but they were written for his invented personages to speak. In Shakespeare's artistic development we can trace his gradual ability to differentiate characters in terms of language. He creates differing registers of speech, and variations in medium. Clowns speak prose, varied with rhymes and songs; servants or socially subordinate characters speak prose; their superiors speak in pentameter verse to each other, but can speak prose to their servants.²¹ In the mature plays we can differentiate up to half-a-dozen characters, each stylistically different. In *Hamlet* we could itemize the linguistic features that individualize the Prince, Claudius, Gertrude, Polonius, Ophelia, and the Gravedigger. In *King Lear*, Lear himself goes through three significant changes of style across the play: authoritative, self-important and angry in Acts 1 and 2; angry, confused and increasingly demented in Acts 3 and 4; and with an utterly chastened style after he recovers sanity, using simple declarative sentences, short words, and a private, not public, tone. Other characters change their styles when they wish according to their altering circumstances. Within a few hundred lines Goneril and Regan mark the change from public performance (where they exude unctuous flattery in verse) by moving to the prose of their private world of ruthless egoism. Edmund (like Iago) introduces himself in verse, but can feign duteous virtue at will. He shifts to prose to share Gloucester's astrological credulity and pessimism, acts the serviceable lover with Regan, and reveals his private plans and hopes in verse soliloquies of great energy and ruthlessness. Edgar is at first the victim of the hypocrite's manipulations, simply trusting, reactive. But when he is hunted like an animal he adopts a linguistic disguise as a 'Bedlam beggar',²² in a jumble of apparently disconnected speech fragments, some of them borrowed from Samuel Harsnet's *Declaration of Egregious Popish Impostures* (1603). But Edgar retains his verse style and his

²¹ Ibid.

²² According to contemporary accounts, this class of beggar feigned madness, claiming to have been imprisoned in Bethlehem Hospital, an asylum for the insane.

true *persona* in his compassionate response to the sufferings of Lear and Gloucester, delivered in asides and soliloquies as he steps in and out of his role as Poor Tom. Confronted by Oswald, seeking to claim the price on his head, Edgar adopts a dialect, as much to act the role of a dumb rustic as to disguise his identity from his blinded father. When he can finally challenge and defeat Edmund and regain his rightful place in the kingdom, he speaks in a style we have not heard before, dignified and forceful. Under the pressure of events Kent also changes his style after being banished by Lear, adopting a bluff combative satirical *persona* in his encounter with Oswald and Cornwall. Gloucester's language also undergoes a transformation, from dutiful attendance in performing his office to a crushed state of nihilism and despair following his blinding and his realization that Edmund has deceived and betrayed him. Only Cordelia and the Fool retain a recognizable style throughout.

Given the presence of nine distinctive idiolects in this play, seven of which display either voluntary or involuntary transformations as the plot develops, what can the function words tell us? To separate them out and subject them to sophisticated statistical manipulation is to perform a self-validating exercise at a considerable distance from the play. By pooling them together, exponents of statistical methods break down any chance of preserving Shakespeare's careful distinction between characters. Having lumped them all together in one pot, they claim that this data allows them to identify Shakespeare's idiolect. This is self-evidently impossible. In *All's Well that Ends Well*, the King rebukes Bertram for having rejected Helena as a wife on the grounds of her inferior social standing. He reworks the old proverb 'there is no difference of blood in a basin':

Strange is it that our bloods,
Of colour, weight, and heat, pour'd all together,
Would quite confound distinction, yet stands off
In differences so mighty.²³

²³ *All's Well that Ends Well*, 2.3.118–21. In his Arden edition (London, 1959), G.K. Hunter offered this paraphrase: 'it is strange that no one would be able to distinguish noble blood from base blood, in respect of

In extracting all the function-words for separate computation, scholars would also ‘confound’ – that is, confuse distinction.

The proponents of function-word counting have never paused to consider whether or not it is an appropriate method to be used in studying drama. Hugh Craig claims that ‘computational stylistics’²⁴ shows that ‘writers leave subtle and persistent traces of a distinctive style through all levels of their syntax and lexis’, making ‘individual selections that persist across all their uses’ of ‘the given language’ (2009: xvi). That is a grandiose claim, and it would be interesting to know how Craig would establish it. But he at once makes an even larger claim, concerning drama. He states that, ‘within the dialogue of Shakespeare and his contemporaries’, the author’s ‘distinctive style’ can be confidently traced:

It persists even when the dramatists strive to create their own fictional linguistic individualities in characters. Hal, Falstaff, and Hotspur do have their own languages but *underlying* them all is a Shakespearean idiom, which means they are all distinct from Jonson, Marlowe or Middleton characters. This idiom is so powerful and persistent that even the computer can detect it. (1989: xvi; my italics)

The concluding assertion may seem paradoxical, for what, after all, can the computer ‘detect’? Only material in the form that its user submits for processing. Craig at once rephrases the point to argue that ‘it wasn’t until the computer came along that we could properly appreciate some aspects of this miraculous secret working of language in these very familiar plays and characters’. But there is nothing either miraculous or secret about the working of language: the process is open and public, otherwise we could never communicate with each other. What Craig means by ‘secret’ is merely the results of computer-driven statistical analyses that were previously invisible to human perception. However, these

colour, weight, or heat, if these bloods were poured together, and yet you think your blood (pedigree) separated by such a difference from hers’ (p. 56 note).

²⁴ Understandably, Burrows and Craig prefer the term ‘computational stylistics’ to ‘stylometry’, associated with the discredited methods of A. Q. Morton. But to count the frequency of occurrence of common words cannot be deemed ‘stylistics’.

minute differences in the frequency with which *the* or *and* occur in a text cannot represent a ‘Shakespearean idiom’ as an absolute ascertainable linguistic fact or series of features. First, they are minutiae of usage, of no appreciable significance; and secondly, they exist below the threshold of our unaided perception. Perhaps Craig implies that the statistical data that result from computation of word-frequencies can differentiate Shakespeare’s characters from those of Jonson, Marlowe, or Middleton? But those are already visible to the naked eye, on the page or in the theatre, and can be immediately experienced. The function word method only covers analyses achieved in a language laboratory.

Craig’s approach throughout is resolutely author-based, and thus anti-dramatic. It ignores the fundamental nature of drama as an *agon* or action that pitches against each other characters having clearly-defined and irreconcilable goals. These conflicts lead to resolutions that involve death in tragedy, a fresh contract or re-ordering of society in comedy, and in historical plays a further demonstration of the constraints on human behaviour within large historical processes. Craig’s single-minded focus is on Shakespeare as an individual author who experiences events and sensations in the real world, as we do. In his terms, ‘since Shakespeare processes the idea or the object in his own way, the expression will bear some stamp of individuality, too; and a scene or an act will become uniquely identifiable’ (4). It may be true that each individual ‘processes’ their experience of the real world differently, but when Shakespeare represents the idea or object in a play it is through the eyes and speech of invented characters, so it is no longer identifiable as his own. Lear sees the world differently in the course of the play, and his view of events also shows his system of values, as it undergoes a massive transformation from concern with himself to an awareness of others. Goneril and Regan process ‘the idea or object’ in their own ways, conflicting with Lear once he has conferred power on them, and ultimately conflicting with each other as they become rivals for Edmund’s love. Although Shakespeare created all three characters, and hundreds

more, his supreme skill as a dramatist was to position himself within each character and to realize how each of them ‘experiences events and sensations’ in their own way, in attempting to realize their own goals. Craig’s conception of characters is static, whereas drama is in all respects dynamic, acting out the forces that affect, and change characters. According to Craig, ‘the computer now allows us to establish the identifiable, distinguishing use of language’ of Shakespeare, ‘foremost’ among Renaissance English playwrights for his ‘consistent style’ (5). This is a deeply mistaken concept. In any case, ‘style’ includes a multitude of linguistic features: tone, register, syntax, range of vocabulary, use of metaphor and other rhetorical figures, many of which cannot be quantified – or if they can, yield meaningless results (such as the number of metaphors per thousand words). The analyses that Craig and his associates perform in the chapters that follow involve a machine-driven reduction of style to word frequencies, computed on a corpus of two hundred function words, which will supposedly identify Marlowe as co-author of *2 Henry VI*,²⁵ and validate the theory that the Quarto and Folio texts of *King Lear* are separate compositions.²⁶

Readers of these and other studies using ‘computational stylistics’ may be impressed by the professionalism with which the battery of statistical tests has been devised and executed. They may also be bemused at the discrepancy between linguistic minutiae and the theoretical structure they are called on to support. Although ‘an author can limit his style, vary it, imitate someone else’, Craig writes, ‘tests of common words, rare words, and word pairings, especially when used in conjunction, can detect the similarities that continue to ride as a foundation *underneath* such easily perceived changes (9–10). Thanks to ‘the software we use for word-counting’ (xvii), he claims, computational stylistics can perceive ‘consistencies’ within the styles of ‘early Henry James and late Henry James’ that would elude ‘even the most perceptive, informed, and experienced readers’ (10). It may well be that James

²⁵ See Hugh Craig, ‘The three parts of *Henry VI*’, *Mystery of Authorship*, 40-77.

²⁶ See Arthur F. Kinney, ‘Transforming *King Lear*’, *ibid.* 181-201. For a radically different account of the interdependence of Quarto and Folio see Brian Vickers, *The One ‘King Lear’* (Cambridge, MA, 2016).

continued to use some word-choices throughout his career (although it would be a travesty to reduce his complex idiosyncratic style to such evidence),²⁷ but what holds for a nineteenth-century novelist does not hold for Renaissance English dramatists. Craig seems determined not to yield this point:

Shakespeare's characters, given different lexicons as a means of characterization, can also be detected as the work of the same author, since such a range of word choices and uses remains in a kind of *envelope* of style that is demonstrably Shakespeare's. Such an 'envelope' still distinguishes Shakespeare from Jonson – or, for that matter, from Marlowe, Middleton, Webster, and Fletcher. (10; my italics)²⁸

Craig now varies his metaphor. Previously the claimed 'consistent style' of Shakespeare was said to be '*underlying*' or riding '*underneath*' what could be experienced on the page or in the theatre. Now it is a mysterious 'envelope', invisible to everyone but the computational stylistician. I can only react to such excessive claims with disbelief. The fundamental fact ignored in his discussion is that there is no such thing as a 'Shakespearean idiom'. Apart from the dedications to his patron in the two narrative poems, every word in the plays is spoken by an invented personage. Some lesser characters are individualized by verbal tics, such as the drawer Francis with his 'Anon, anon sir', or Nym with his 'That's the humour of it'. Others are given senile repetitions (Polonius, Justice Shallow) or 'foreigners' English' (Dr Caius in *Merry Wives*, Armado, Fluellen). Of the characters important to the dramatic action, many are differentiated by distinctive styles, embracing lexicon, grammar, syntax, rhetorical figures and tropes, and – guiding everything – the good or evil purposes to which these resources are put. Not a single word out of this vast corpus can be identified with Shakespeare – indeed, it would be a major failing for any dramatist if this were true of their work. Dramatists write words for characters to speak.

²⁷ See, e.g., David L. Hoover, 'Corpus stylistics, stylometry, and the styles of Henry James', *Style* 41 (2007): 174–2003.

²⁸ It must be noted that, although Craig is undoubtedly right to state that Shakespeare gives his characters 'different lexicons', his method of computing only high-frequency words prevents him from studying them.

II

It may be the case that the computer-driven statistical analysis of function words can distinguish between different dramatists on the basis of works whose authorship is known, and that is no doubt an achievement of some kind, given that common words provide the framework for larger patterns of meaningful utterance. But some of the analyses performed using this method have rejected accepted authorship attributions dating back to the First Folio or even earlier. In the volume co-authored by Craig and Kinney, Craig applied function-word computation to the three *Henry VI* plays, in order to address what he calls ‘perhaps the thorniest problem in attribution to the Shakespeare canon’.²⁹ Craig first divided the plays into 2,000-word segments, then examined ‘how much the segments departed from *Shakespeare norms*’ according to the outcome of two tests. One of them used function words, the other lexical words, and the results were evaluated by reference to ‘the 27 plays making up the core Shakespeare canon’ (45; my italics). But from the outset Craig overlooked the crucial importance of chronology. All serious scholars date *1 Henry VI* as later than the other two plays, with Martin Wiggins offering this dating: *Part 2*, 1591; *Part 3*, 1591; *Part 1*, 1592. In other words, Part 2 was Shakespeare’s first play, and if tests show that it does not meet ‘Shakespeare norms’ – however defined – which have been set by later plays, then the obvious explanation is that Shakespeare was as yet in the process of establishing norms of usage at every level of his art, from dramaturgy to characterization to all aspects of language. Craig, with his single-minded concern with ‘the question of how distinguishable or otherwise is one author’s style in a field like Shakespearean drama’ (47), never grasps this point. Irregularities in the frequency counts for function and lexical word tests, he decides, present ‘a quite different pattern from the canonical plays ..., a pattern that suggests mixed authorship’ (48). Reviewing the tests showing that ‘some *1* and *2 Henry VI* segments stray

²⁹ ‘The three parts of *Henry VI*’, in *Mystery of Authorship*, pp. 40–77 (40).

outside' Shakespeare norms, Craig concludes that 'the most obvious explanation is that this is the result of mixed authorship' (50).

Here, already, Craig's method reveals its limitations. Although his discussion is confused and inconclusive,³⁰ he is correct in treating *1 Henry VI* as a collaborative play, for scholars since the eighteenth century have commented on its great variety of styles. *2 Henry VI*, by contrast, has been almost universally accepted as Shakespearean throughout. It is unfortunate that Craig, having ignored Dover Wilson's admirable identification of three Shakespearean scenes in *1 Henry VI*, should take seriously Wilson's unjustified division of *2 Henry VI* between Greene, Peele, Marlowe, and Nashe. Craig also accepted the doubtful arguments of Paul Vincent that 'there were at least two different authors in the play', on the slender grounds that the Folio text has alternative spellings of 'o' and 'oh', and that the 'dearth of the pronoun *ye* and peculiarities in classical allusions made Act III distinct from the rest of the play and associates it with Shakespeare'.³¹ The proper scholarly response to this dubious claim would be to ask whether these supposed deviations from a putative 'Shakespeare norm' in this his first play are significant; and if so, what other explanations

³⁰ See pp. 51–68, where Craig reviews previous authorship attributions, not always correctly reported. It is not true that John Dover Wilson, in his edition *The First Part of King Henry VI* (Cambridge, 1952), agreed with Malone in attributing the whole sequence presenting the siege of Bordeaux (4.2–4.7.32) to Shakespeare. Wilson argued forcefully, and with a sensitivity to changes in verse style rare among current attribution scholars, that 4.5 and 4.6, both of which present the last encounter between Talbot and his son before the fatal battle, are duplicates. As he judged, 4.6 is 'virtually a repetition' of 4.5, an identical action with the same arguments and at times the same words, but 'of very different poetic quality'. Wilson showed that the couplet rhymes in 4.5 are end-stopped, Shakespeare's normal practice, while those in 4.6 'are looser, often overrun' the verse line, and use a 'poverty-stricken diction'. He attributed 4.5 to Shakespeare, 4.6 to Greene, his usual suspect, an attribution that we need not accept. But Wilson is to be congratulated for making the indisputable point, for the first time in the centuries-long discussion of this play, that one scene 'was written to replace the other', and that '4.5 is Shakespeare's revision of 4.6' (op. cit., xliii–xlvi, 182–6). In discussing Wilson's ascription of Act 1 to Nashe, now universally accepted, Craig once again ignores the question of genre, comparing that tragic history with Nashe's only other dramatic work, the moral masque, *Summer's Last Will and Testament*, a quite different type of play. He then compares them both with Nashe's satirical prose works, a different genre again (54–6). This is to ignore the accepted principles of modern attribution studies. For a different account, see Brian Vickers 'Incomplete Shakespeare: or, denying co-authorship in *1 Henry VI*', *Shakespeare Quarterly*, 58 (2007): 311–52; for the identification of Kyd as author of Acts 2–5 of the original play, described by Henslowe as '*harey the vj*' when it was performed by Lord Strange's Men at the Rose theatre in June 1592 see Brian Vickers, 'Thomas Kyd, secret sharer', *Times Literary Supplement*, 18 April 2008, pp. 13–15. The Chamberlain's Men acquired the play in 1594, following the demise of Strange's company, and at some point before 1598 Shakespeare added what we now know as *1 Henry VI* 2.4, 4.2, and 4.5.

³¹ See P. J. Vincent, 'Unsolved mysteries in *Henry VI, Part Two*', *Notes and Queries* 48 (2002): 270–4.

there might be. But Craig unquestioningly accepts it as supporting ‘the idea that the play is a collaboration’ (69). Unfortunately, the Folio Act divisions, as used by Vincent, do not correlate with the 2,000-word segments into which Craig has divided the play. According to his word-frequency counting, three ‘segments’, numbers 6, 7 and 8, ‘which emerge as the *most* “Shakespearean” ... include all but the first 100 or so lines of Act III’ (ibid.; my italics). Whoever consults the play will discover that this sequence (lines 1–113), belonging to Craig’s ‘segment 5’, is a crucial part of the scene. The remaining 270 lines work through the consequences of its opening, the quarrel in parliament between the good Duke Humphrey and his enemies, Queen Margaret, Suffolk, and Richard Duke of York, with the King powerless to intervene. On Craig’s segmental division these 113 lines would be among the *least* ‘Shakespearean’. His ‘segment 8’ ends at 4.1.36, well into the prolonged murder of Suffolk, which concludes with the victim’s severed head and body being brought on stage, a ‘barbarous and bloody spectacle’ indeed (4.1.145). Are these 130 lines also less Shakespearean?

The lack of fit between Craig’s 2,000-word segments and the Act divisions with which he tries to correlate them is awkwardly evident as he proceeds to the most contentious part of his essay. He has nothing to say about ‘segment 9’, which covers 4.1.36 to 4.2.60, but he reports that ‘segments 10 and 11’, which stretch from 4.2.160 to 5.1.13 are, according to his test, ‘clustered with the Marlowe segments’, from which he makes this inference:

All of the complete scenes that are included involve what the 1594 Quarto title-page calls ‘the notable rebellion of *Jack Cade*.’ Is it possible that the Cade scenes are by Marlowe rather than Shakespeare? Certainly they are detachable from the rest of the play. Some characters in them are shared with other sections – the King, Queen Margaret, and Buckingham in particular, but the majority appear only in 4.2–9... (70)

This inference must be challenged on several grounds. First, his ‘segment 10’ begins at 4.2.160, well past the scene’s crucial beginning, with its stage-setting: ‘*Enter two of the*

Rebels with long staves'.³² Shortly afterwards Jack Cade and his followers enter, also '*with long staves*', and the comical-tragical-anarchical rebellion proceeds. The Clerk of Chatham is dragged on, a suspicious figure: 'he can write and read and cast account' – 'O monstrous!', Cade exclaims, and soon sends the innocent man to his death: 'Hang him with his pen and inkhorn about his neck' (4.2.109–10).³³ Next, Cade knights himself in order to be socially 'equal' with Sir Humphrey Stafford, the first representative of authority to confront the rebels. The rebels announce that they intend to 'have the Lord Say's head for selling the dukedom of Maine' and, more serious still, because 'he can speak French, and therefore he is a traitor'. Stafford exclaims 'O gross and miserable ignorance!' (4.2.168) just before this 'segment' ends. No one can deny that this sequence, including all but twenty lines, is an integral part of the scene. Yet Craig reports that, on his computations 4.2 was 'only classified as Marlowe by one of the lexical-words tests ... This is a little perplexing' (73). Indeed it is. This discrepancy might have caused Craig to ask himself whether his method was truly reliable, or whether there might be some other explanation – such as the fact that Shakespeare was synthesizing historical material from Edward Hall's *Union of the Two Noble and Illustre Famelies of Lancastre and Yorke* (1548) with accounts of the Peasant's Revolt from Grafton's *Chronicle at large* (1569) and Holinshed's *Chronicles* (1587).³⁴ The absorption of source material running to several thousand words must have had an influence on Shakespeare's lexicon (a topic that computer-assisted research could well explore), and should have been considered.

The second necessary objection to Craig's inference of Marlowe's authorship is his statement that 'Certainly [these scenes] are detachable from the rest of the play'. In fact,

³² I quote from *The First Part of the Contention 1594*, ed. William Montgomery (Oxford, 1985; Malone Society Reprints), which probably represents how the play was performed in 1591-2.

³³ Shakespeare's presentation of anarchy as being hostile to literacy has been justified in our times, as in the despotic rule of Pol Pot in Cambodia.

³⁴ See, e.g., Geoffrey Bullough (ed.), *Narrative and Dramatic Sources of Shakespeare, Volume Three, Earlier English History Plays: Henry VI, Richard III, Richard II* (London and New York, 1960), pp. 95–8, 113–18, 128–33.

Shakespeare carefully prepared for Cade's rebellion in the long soliloquy by York ending 3.1 (a 'segment' that Craig earlier attributed to Shakespeare). Having been instructed to put down the Irish rebellion, York sees his chance to foment disorder in England:

And, for a minister of my intent,
I have seduc'd a headstrong Kentishman,
John Cade of Ashford,
To make commotion, as full well he can,
Under the title of John Mortimer.
In Ireland have I seen this stubborn Cade
Oppose himself against a troop of kerns,
And fought so long, till that his thighs with darts
Were almost like a sharp-quill'd porpentine; (3.1.355)

In addition to his prowess in battle, York reveals that Cade has acted as a double agent, having

Full often conversed with the enemy,
And undiscover'd come to me again
And given me notice of their villainies.
This devil here shall be my substitute;
For that John Mortimer, which now is dead,
In face, in gait, in speech he doth resemble. (368–73)

Although this preparation hardly squares with the character of Cade that we see in Act 4, Shakespeare took this (unhistorical) alliance from Hall's chronicle, who described a well-prepared plot for 'some great commocion and rising of people to be made against the King'.³⁵ Shakespeare duly dramatizes the main stages of the Yorkist inspired rebellion, the murder of Stafford and his brother, their seizure of London Bridge, killing of Matthew Goffe, beheading of Lord Say, and other random atrocities. These scenes are not 'detachable' from the play, since they dramatize the accelerating decline of England into chaos.³⁶

³⁵ Hall, ed. cit., pp. 113–17.

³⁶ See the classic essay by J. P. Brockbank, 'The frame of disorder – *Henry VI*', in J. R. Brown and Bernard Harris (eds.) *Early Shakespeare* (London, 1961), pp. 73–99, reprinted in Philip Brockbank, *On Shakespeare: Jesus, Shakespeare and Karl Marx, and Other Essays* (Oxford, 1989), pp. 79–103.

In the source, order is restored when the King issues ‘a general pardon of all the offenders’, who secretly slink away ‘without bidding farewell to their captain’. Shakespeare avoids this anti-climax by adding a confrontation between Cade and the aged hero, Clifford, whose invocation of England’s hero, Henry V, quells the uprising, leaving Cade to lament the fickleness of ‘this multitude’ (4.8.55–6). Cade’s rebellion thus forms an integral part of the play even after he has been killed by ‘Alexander Iden, an esquire of Kent’ (4.10.43), for in the following scene York returns with ‘his army of Irish’ (5.1.0), looking to profit from the discord he has created. According to Hugh Craig’s segmentation, only the first 13 lines of this scene fitted into his Marlowe ‘segment’, leaving us wondering whether he imagines that Shakespeare, or someone else, took up the pen at this point to write the remaining 337 lines of the play, which ends with ‘a glorious day, Saint Albans battle won by famous York’ (5.3.29–30). In retrospect it is evident that Cade’s rebellion was an integral part of York’s plot, and Shakespeare’s, and cannot be deemed ‘detached’.

When Hugh Craig found it ‘a little perplexing’ that 4.2 did not have the same ‘likeness to Marlowe’ that his tests detected elsewhere, he could have questioned his own procedures; instead he pressed on with the Marlowe connection. As we have seen, the Cade rebellion begins 160 lines earlier, at the start of 4.2, but Craig found that this portion of the play ‘is only classified as Marlowe by one of the lexical word-tests’, not sufficient support by his criteria. He acknowledged that it fails to show – as he puts it – ‘the likeness of Marlowe in style and vocabulary that is so strong’ in the rest of these scenes. Having reached this impasse, Craig fell back on impressionistic literary criticism. He began by looking for similarities between Cade and Joan la Pucelle in *1 Henry VI*, whom he had earlier decided was created by Marlowe:

However, Marlowe clearly has a large hand in the latter part of the Cade sequence. Cade is plausible as a Marlowe character. Like Joan in *1 Henry VI*, he is of humble parentage but aspiring to supreme political power. This is a Marlovian theme, familiar from the shepherd Tamburlaine ... Both Joan and Cade construct fantasies of noble

birth ... Cade at his death continues to spout defiance like a good Marlovian hero...
(73).

Anyone familiar with this period might observe that there are many characters in Elizabethan drama who aspire 'to supreme political power' or 'spout defiance' on being captured without therefore being declared 'a good Marlovian hero'. But few, if any, of these aspiring rulers are presented in prose, a significant fact, especially in connection with Marlowe. Had Craig considered other explanations he might have remembered that Marlowe wrote very little prose. When he did, he followed one convention of Elizabethan drama, whereby a character's descent into madness is represented by a descent from verse to prose. So, in *Tamburlaine*, Zabina, having discovered her husband's corpse, 'raves' in prose, as some editors indicate, before killing herself.³⁷ Marlowe also used prose for internal contrasts within a play, but he never wrote scenes constituting several hundred lines of prose. *Dr Faustus* might seem a partial exception, but modern scholarship has identified the comic scenes in both the *A* and *B* versions as the work of other hands, possibly Henry Porter and William Birde.³⁸

A further aspect of the Cade scenes that Hugh Craig overlooked is that he is unlike a Marlovian or any other kind of hero.³⁹ In the source Cade is a threatening figure. Hall describes him at first as a 'capitayn', a detail that may have given Shakespeare a hint for the prowess in Ireland that York attributes to him, but the historian was using irony or sarcasm. Hall subsequently refers to 'the Kentish capitayn, or the covetous Cade' (the same person, with mocking alliteration), and with increasing disapproval: 'this glorious Capitayn' is 'subtill' (deceiving), and surrounded by 'a multitude of evil rude and rusticall persones'. The King's envoys found him to be 'arrogant in hart', disliking 'the wilful pertinacity, and manifest contumacie of this rebellious Javelyn ... thys outrageous losel'. Hall's narrative

³⁷ See *1 Tamburlaine*, 5.1.305–19.

³⁸ See David Bevington and Eric Rasmussen (eds.), *Doctor Faustus. A- and B-Texts (1604, 1616)* (Manchester and New York, 1993), especially pp. 62–73.

³⁹ The nearest character to a Marlovian hero in *2 Henry VI* is of course the aspiring Machiavellian York.

continues with Cade's atrocities: he has Lord Say killed before he has finished his final prayers on the scaffold.⁴⁰ Even more outrageous is Cade's treatment of James Cromer, Say's son-in-law: '& hym without confession or excuse heard, caused there lykewyse to be hedded, and his head to be fixed on a poole, and with these two heddes this bloody butcher entered into the citie agayn, and in despite caused them in every strete, kysse together, to the great detestacion of all the beholders...' (cf. 4.7.109–12, 130–1).

Shakespeare includes all of these horrible acts in the play and he could easily have portrayed Cade as a frightening figure. But as he read on in Hall, immediately after that passage, he came across an aspect of Cade's behaviour that caused him to avoid such an easy choice. Cade feared that he might be exposed as a low-born imposter:

He also put to execucion in Soutwarke divers persons, some for infryngyng his rules and precepts ... other he tormented of his olde acquayntance, lest they shoulde blase & declare his base byrthe, and lowsy lynage, disparagyng him from his usurped surname of Mortymer...

Geoffrey Bullough helpfully noted an additional passage in Fabyan where he 'describes the murder of one Baylly who "was of the familier and olde acquaintance of Jake Cade, wherfore so sone as he espyd him comming to himward, he caste in his minde that he would discover his living and olde maners, and shew of his vile kinne and lynage". So he had Baylly slain as "an enchaunter and of evyll disposicion".'⁴¹ These references in the sources gave Shakespeare the lead in turning this 'caitiff' into both a frightening and a comic figure, lacking legitimacy, a leader who is mocked by his own supporters.⁴² In his first appearance Cade's entourage includes Dick the butcher and Smith the weaver, to whom Shakespeare gives a series of asides debunking their leader's claims to noble birth:

CADE My father was a Mortimer –

⁴⁰ The Ghost of Hamlet's father describes how he had been, 'Cut off even in the blossoms of my sin, | Unhous'led, disappointed unanel'd', denied the Eucharist and the last unction (1.5.76–9).

⁴¹ Bullough, p. 116 note, citing Robert Fabyan, *The New Chronicles of England and of France*, ed. H. Ellis (London, 1811), p. 451.

⁴² See Vickers, *Artistry of Shakespeare's Prose*, pp. 9–10, 26–7.

DICK [*Aside*] He was an honest man, and a good bricklayer.
CADE My mother a Plantagenet –
DICK [*Aside*] I knew her well; she was a midwife.
CADE My wife descended of the Lacies –
DICK [*Aside*] She was indeed a pedlar's daughter, and sold many laces.
SMITH [*Aside*] But now of late, not able to travel with her furr'd pack, she
washes bucks here at home.
CADE Therefore am I of an honorable house—
DICK [*Aside*] Ay, by my faith, the field is honorable, and there was he born,
under a hedge; for his father had never a house but the cage.
(4.2.39–52)

Shakespeare used the same technique of the deflating aside in other plays. The conspirators in *Twelfth Night* mock Malvolio's posturing from their safe hiding place (2.5.15–185). In *All's Well that Ends Well* the conspirators get the blindfolded Parolles to give his candid and scabrous verdict on Bertram, who is listening to this 'Damnable both-sides rogue', giving a running commentary in asides (4.3.96–329). Shakespeare used this comic device for the final time some twenty years after *2 Henry VI*, in *Cymbeline*, also at a character's first appearance. Here it is the boor Cloten, attended by two Lords, having just completed some kind of fencing match:⁴³

CLOTEN Have I hurt him?
SECOND LORD [*Aside*] No, faith; not so much as his patience.
FIRST LORD Hurt him? His body's a passable carcass, if he be not hurt. It is a
thoroughfare for steel, if it be not hurt.
SECOND LORD [*Aside*] His steel was in debt, it went o'th'backside the town.
CLOTEN The villain would not stand me.
SECOND LORD [*Aside*] No, but he fled forward still, toward your face.
FIRST LORD Stand you? You have land enough of your own, but he added to
your having, gave you some ground.
SECOND LORD [*Aside*] As many inches as you have oceans. Puppies!
CLOTEN I would they had not come between us.
SECOND LORD (*aside*) So would I, till you had measured how long a fool you
were upon the ground.
(1.2.5–24)

⁴³ Ibid., 409–11.

Unusually, Shakespeare repeats this dramatic effect when we next see Cloten (2.1.1–51). Both butts also condemn themselves out of their own mouths. Cade is given a soliloquy in which his pretensions are comically displayed (4.10.1–15); Shakespeare does the same for Cloten, twice (3.5.130–45; 4.1.1–25). In both cases the pretender to a higher station is soon killed by a virtuous character: Alexander Iden in Kent, Guiderius in the Welsh hills.

Hugh Craig's use of both function and lexical words to question Shakespeare's sole authorship of *2 Henry VI* is a failure on every count.⁴⁴ Despite the diligence with which he carried out his tests, and the care with which he explained the statistical procedures, only one conclusion is possible: this atomistic approach to language, using only function words, ignoring the difference between authors and their characters, is wholly unsuitable for drama.

II

In creating his plays Shakespeare explored the full expressive potential of language. Indeed, he reached a degree of imaginative fusion of all the elements of style – verse form, lexis, syntax – that has never been equalled in English drama. In order to do justice to his achievement, or that of his fellow dramatists, it seems obvious that scholars should possess a theory of language, or a linguistic awareness, appropriate to their task. The fundamental principle is that language is sequential.

It has long been recognized that natural languages derive their ability to communicate meaning from the uniting, composing, or 'weaving together' of their constituent elements. The metaphor of weaving is at least as old as Plato, who used it in two dialogues, as Deborah Gera has noted.⁴⁵ In the *Statesman* (277d–278h) 'Plato speaks of the interweaving (*sumplokē*) of letters into syllables: he describes how young children are taught to recognize the vowels and consonants which are interlaced to form syllables and words'. From the phonological

⁴⁴ For a telling critique, based on different grounds, see Darren Freebury-Jones, 'Exploring co-authorship in *2 Henry VI*', *Journal of Early Modern Studies* 5 (2016): 201–16.

⁴⁵ Deborah Levine Gera, *Ancient Greek Ideas on Speech, Language, and Civilization* (Oxford, 2003), p. 206.

level of language Plato used the weaving metaphor in the *Sophist* to describe the creation of a meaningful utterance (in this context *logos* means ‘statement’). The Eleatic Stranger leads Theaetetus through a sequence of reasoning in which (if I may record only one part of the dialogue) he accepts that

The signs we use in speech to signify being are of two kinds, one called ‘nouns’ (*onomata*), the other ‘verbs’ (*rhēmata*). By ‘verb’ we mean an expression which is applied to actions, and by a ‘noun’ [or ‘name’] the spoken sign applied to what performs these actions. Now a statement never consists solely of nouns spoken in succession, nor yet of verbs apart from nouns: these words spoken in a string this way do not make a statement.

To explain his argument, the Eleatic Stranger cites a sequence of ‘verbs signifying actions’, such as ‘walks runs sleeps’ which can be uttered one after the other without amounting to a statement.

And again, if you say ‘lion stag horse’ and any other names given to things that perform actions, such a string never makes up a statement. Neither in this example nor in the other do the sounds uttered signify any action performed or not performed or nature of anything that exists or does not exist, until you *combine* verbs with nouns. The moment you do that, they *fit together* and the simplest combination becomes a statement of what might be called the simplest and briefest kind. For instance, when one says ‘A man understands’, this is a statement of the simplest and shortest possible kind, because now it gives information about facts or events in the present or past or future; it does not merely name something but gets you somewhere *by weaving together* verbs with nouns. Hence we say it ‘states’ something, not merely ‘names’ something, and in fact it is this complex that we mean by the word ‘statement.’⁴⁶

Although only an analogy, as Deborah Gera comments, ‘here we encounter a view of language as something more complex than a concatenation of names’ (206).

This concept of language as a coherent sequence recurs in the history of linguistics. In the work of Ferdinand de Saussure (1857–1913), the acknowledged founder of modern linguistics, it is expressed through the metaphor of a chain. In his great work, the

⁴⁶ *Sophist*, 261e–262d, in the translation by F.M. Cornford, reprinted in *Plato. The Collected Dialogues*, ed. Edith Hamilton and Huntington Cairns (New York, 1963), pp. 1008–9; my emphases. Students of Greek philosophy will know that Plato applies these analogies with weaving and the components of language to a complex sequence of argument: for helpful commentary see, e.g., I. M. Crombie, *An Examination of Plato’s Doctrines. Volume II. Plato on Knowledge and Reality* (London, 1963), pp. 115, 374–88, 411–16, 492–8.

posthumously-published set of students' lecture notes known as the *Cours de linguistique générale* (1916),⁴⁷ Saussure introduced the notion of 'language articulation', derived from the Latin word *articulus*, which 'means "member, part, subdivision in a sequence of things" [and] may refer to the division of the chain of speech into syllables, or to the division of the chain of meanings into meaningful units' (*CLG*, p. 27). Saussure's metaphor of 'the chain of speech' and 'the chain of meaning' expresses his recognition of the linearity of language, as an event in time: 'the linguistic signal, being auditory in nature, has a temporal aspect, and hence certain temporal characteristics: (a) it is linear: *it occupies a certain temporal space*, and (b) *this space is measured in just one dimension: it is a line*' (p. 103; author's emphases).

This seldom recognised principle is basic to language:

The whole mechanism of linguistic structure depends upon it. Unlike visual signals (e.g. ships' flags) which can exploit more than one dimension simultaneously, auditory signals have available to them only the linearity of time. The elements of such signals are presented one after another: they form a chain. This feature appears immediately when they are represented in writing, and a spatial line of graphic signs is substituted for a succession of sounds in time. (*ibid.*)

Saussure used various terms to describe the 'linearity' of the sign, the most influential being based on the Greek word *syntagma* ('that which is put together in order'):

Words as used in discourse, strung together one after another, enter into relations based on the linear character of languages. Linearity precludes the possibility of uttering two words simultaneously. They must be arranged consecutively in spoken sequence. Combinations based on sequentiality may be called *syntagmas*. ... (170).

Saussure emphasized that such 'syntagmatic relations hold *in praesentia* ... between two or more terms co-present in a sequence' (171). It follows that 'the notion of a syntagma applies

⁴⁷ After Saussure's death Charles Bally and Albert Sechehaye published the *Cours de linguistique générale* (Paris, 1916), from the students' notes of his lectures given at the University of Geneva between 1907 and 1911. Fresh examination of the surviving students' notes in the 1950s revealed that the original editors had imposed their own arrangement on this material and, in several places, added their own wording. It was not until 1996, when many of Saussure's own manuscripts, dating back to 1891, were discovered in the Orangery of the Saussure family home in Geneva, that the speculative nature of their reconstruction became clear. I quote from his *Cours de linguistique générale* in Tullio de Mauro's 'Edition critique' (Paris, 1972, 1985), and in the translation by Roy Harris, *Course in General Linguistics* (London, 1983).

not only to words but to groups of words, and to complex units of every size and kind (compound words, derivative forms, phrases, sentences)' (172).⁴⁸

Zellig Harris, an influential American linguist whose long career spanned many developments in linguistic theory, took it as axiomatic that

The parts of a language do not occur arbitrarily relative to each other: each element occurs in certain positions relative to certain other elements. The perennial man in the street believes that when he speaks he freely puts together whatever elements have the meaning he intends: but he does so only by choosing members of those classes that regularly occur together, and in the order in which these classes occur'.⁴⁹ In speaking or writing we make choices: 'we build up a stock of utterances each of which is a particular combination of particular elements. And this stock of combinations becomes a factor in the way later choices are made ... for language is not merely a bag of words but a tool with particular properties which have been fashioned in the course of its use. (155–6)

It follows that the appropriate unit of linguistic analysis is 'a consecutive (or seriate) discourse' in which we study the 'flow of speech' that contains 'successive elements or sections in discourse', bearing in mind that 'various types of discourse have various types of succession (of sentences, clauses, or other intervals' (158).

The provision of large corpora of actual speech usage in the 1970s, together with the new possibilities of electronic word processing, enabled scholars to create concordances which identified for the first time the enormous role played in natural languages by recurring phrases sometimes known as 'word-strings', another metaphor for the sequentiality of language. These discoveries gave rise to the new sub-discipline of Corpus Linguistics, which has shown that we communicate more often in phrases than in single words. (Linguistics

⁴⁸ In his *Introduction to Theoretical Linguistics* (Cambridge, 1971), John Lyons revived Saussure's metaphor. Discussing 'Concatenation and linearity' in grammatical structures, he observed that 'the combination of words resulting from the application of a grammatical rule' constitutes 'a string, with the order or the concatenated words being determined by the order in which the words occur in sentences of the language' (p. 209). 'That simple sequential model', he added, 'has obvious limitations, for linguists have long been aware that 'sentences are not just linear sequences of elements, but are made up of "layers" of *immediate constituents*, each lower-level constituent being part of a higher-level constituent' (210–11; author's emphases). Nevertheless, linearity applies.

⁴⁹ Zellig S. Harris, 'Distributional structure', *Word*, 10 (1954):146–62 (146).

describe these phrases or collocations as ‘*N*-grams’, where ‘*n*’ is a measure of length: a two-word phrase is a bigram, three words form a trigram and so on.) A study by Bengt Altenberg using the London-Lund Corpus, which contains ‘nearly half a million running words’, found that it included ‘over 201,000 recurrent word-combinations’, amounting to ‘over 80 per cent of the whole’.⁵⁰ Altenberg argued that the ‘recurrent clause element sequences’ in this material ‘can be regarded as a series of overlapping and interlocking options that are utilized again and again by speakers in ongoing discourse’. These sequences, ‘more or less prefabricated or routinized building blocks that are at the speaker’s disposal in the production of discourse’, justify ‘Nattinger’s characterization of speech as “basically a ‘compositional’ process, one of ‘stitching together’ preassembled phrases into discourse”’.⁵¹

Plato’s metaphor of weaving finds a modern equivalent in ‘stitching together’. The discovery that human beings use prefabricated ‘blocks’ or ‘chunks’ of language has been influential in studies of language acquisition, whether in the mother tongue or in learning a foreign language. A recent study treats the following positions as axiomatic: ‘Language is sequential. Speech is a sequence of words. Writing is a sequence of symbols. Learning to understand a language involves parsing the speech stream into chunks which reliably mark meaning.’⁵² These chunks are the ‘patterns’ that occur ‘within the sequence of language’, and ‘at some level of analysis [they] refer to meaning’. Thus ‘language acquisition is essentially a sequence learning problem: the acquisition of word form, collocations, and grammatical class information all result from predominantly unconscious (or implicit) analysis of sequence information ... Phonology, lexis, and syntax develop hierarchically by repeated cycles of differentiation and integration of chunks of sequences’ (41).

⁵⁰ Bengt Altenberg, ‘On the phraseology of spoken English: the evidence of recurrent word-combinations’, in A.P. Cowie (ed.) *On Phraseology. Theory, Analysis, and Applications* (Oxford, 1998), pp. 101–22 (102).

⁵¹ *Ibid.*, p. 115, citing J. R. Nattinger, ‘Some current trends in vocabulary teaching’, in R. Carter and M.J. McCarthy (eds.) *Vocabulary and Language Teaching* (London, 1988), pp. 62–82 (76).

⁵² Nick C. Ellis, ‘Memory for language’, in P. Robinson (ed.), *Cognition and second language instruction* (Cambridge, 2001), pp. 33–8 (40).

If linguists and philosophers of language over two and a half millennia have emphasized the crucial element of composition, fitting words together into sequences that communicate meaning, it is essential that those working in authorship attribution should devise methods that respect this fundamental property of language. Patrick Juola, a foremost scholar in authorship attribution studies, equally well-versed in mathematics and linguistics, recently published a comprehensive survey of the current state of this discipline.⁵³ He observed that

a text is structured as an ordered stream of separate ‘events’ drawn from a population of potential events. These events may be sounds, letters, words, or perhaps even phrases or sentences. Of course, language is not the only system that can be so described: researchers have studied other ‘paralinguistic’ systems (such as music). Furthermore, the relation between the different events in the same stream is not random but governed by high-order regularities’ (252).

Each of these word classes needs the others. Some methods recognize these complexities, including ‘context-free grammars’ and syntactic models, but ‘more often language is treated as “a bag of words”, a collection of every word that appears in the document without regard to order...’ (253). Juola discusses a number of methods, pointing out the need to choose one appropriate to the genre or text-form being studied,⁵⁴ and emphasizing that ‘the best results will come from analysis of an extremely broad set of features, covering many different approaches’ (269). Although rating mathematical and statistical methods highly, Juola underlined the centrality of ‘grammatical and semantic structure’ (255), and the need to use an ‘approach that combines lexical and syntactic information in the use of word *N*-grams (bigrams, trigrams, etc.) to capture words in context’ (265).⁵⁵

The appropriate method, then, for investigating authorship problems in early modern

⁵³ Patrick Juola, ‘Authorship attribution’, *Foundations and Trends in Information Retrieval*, 1 (2006): 233–334.

⁵⁴ Cf. *ibid.*, pp. 246, 286, 297, 317.

⁵⁵ In an ad-hoc authorship attribution competition held in 2004, two of the three most successful methods used *n*-grams (293–4, 295–6). As Juola summed up the results of this competition: ‘Methods that do not use syntax in one form or another, either through the use of word *n*-grams or explicit syntactic coding tend to perform poorly’ (320).

drama will treat words in context, not as comparable elements but as interlocking sequences, by means of which characters produce meaningful utterances. Attribution scholars who use ‘computational stylistics’ are not interested in meaning, for they count the sequences of letters that form a word purely as a separate unit of recurrence. To use Saussure’s terminology, they study the signifier, and ignore the signified. Fortunately, a recent development in electronic word-processing allows us to isolate word strings that convey meaning. The ability of computers to recognize longer sequences of characters has given rise to the production of anti-plagiarism software, designed to expose students who cheated in exams. Using this software, when two texts are compared, the program instantly identifies the recurrence of identical phrases in both, phrases of any length from two words upwards. I use this program not to detect plagiarism but to identify instances where a dramatist repeats phrases, a very common event in early modern drama. Self-repetition is a phenomenon found in literary language of all periods and cultures, between (at least) Homer and Samuel Beckett.⁵⁶ It is particularly common in early modern drama, partly due to the material circumstances of competing theatres, each using repertory system, each keen to emulate the rivals’ successes, with dramatists working with co-authors in turning out anything from two to ten plays a year.⁵⁷

In order to verify the authenticity of *2 Henry VII* I have compared it with *3 Henry VI*, using WCopyfind⁵⁸ to identify every collocation of three words or more found in both plays: there were 606. This is the largest number of matching trigrams I have found over the eight-year period I have been using this method. Using InfoRapid Search & Replace,⁵⁹ I then checked each of these against a database containing all the plays performed in the public

⁵⁶ See, e.g., Martin Mueller, *The Iliad* (2nd edition; London, 2009) and Brian Vickers, ‘Repetition and emphasis in rhetoric theory and practice’, in Andreas Fischer (ed.), *Repetition* (Tübingen, 1994), 85–113, with examples from *3 Henry VI*, Dickens, *Bleak House* (1853), and Beckett, *The Unnameable* (English version, 1958).

⁵⁷ See Brian Vickers (ed.) *The Collected Works of John Ford*, vol. 2 (Oxford, in the press), Introduction: ‘Dramatists repeat themselves’, (00–00).

⁵⁸ <http://plagiarism.bloomfieldmedia.com/wordpress/software/wcopyfind/>

⁵⁹ <http://www.inforapid.de/html/searchreplace.htm>

theatre between 1579 and 1596. I list in Appendix 1 the matches that were unique to these plays, which amounted to 130, another score far higher than anything I have experienced so far. Thus the proportion of specifically Shakespearian collocations is 21.9 per cent; or, to put it in other terms, a general match between the two plays occurs once every five lines, and a uniquely Shakespearian match every 24 lines.

TABLE 1: 2 *Henry VI*

<i>Act, Scene</i>	<i>Lines</i>	<i>Number of matches</i>	<i>Frequency of matches: every x lines</i>
*1.1	259	11	23.6
*1.2	107	4	26.8
1.3	220	11	20.0
1.4	80	3	26.7
2.1	201	8	25.1
2.2	82	2	41
*2.3	105	3	35
2.4	110	8	13.8
3.1	383	19	20.1
3.2	412	16	25.8
3.3	33	4	8.25
4.1	147	1	147
*4.2	190	7	27.1
*4.3	18	1	18
*4.4	60	3	20
*4.5	12	0	0
*4.6	15	1	15
*4.7	136	0	0
*4.8	69	1	69
*4.9	49	3	16.3
*4.10	84	2	42
5.1	216	13	16.4
*5.2	90	2	45
5.3	33	4	8.25
Total	3,111	130	23.9

Asterisks mark scenes that Craig and Burrows judge to be by other hands

The remarkable number of matches establishes conclusively that Shakespeare wrote the whole play, with no other hand intervening. The 13 scenes that Burrows and Craig claim as

unShakespearian contain a total of 38 unique matches for 602 lines, a ratio of one every 16 lines. The remaining 11 scenes that they deem to be Shakespearian, totalling 2,509 lines, have 92 matches, or one every 27 lines, a smaller ratio. Thus their authorship division is self-evidently false.

The greater density of uniquely Shakespearian matches, setting the bar at one match every 28 lines, are 1.1, 1.2, 1.3, 2.1, 2.4, 3.1, 3.2, 4.2 and 5.1. These scenes play out the major public encounters between the Lancastrian and Yorkist camps. The first six of those scenes dramatize the plots by which Humphrey, Duke of Gloucester and uncle to the King, is brought down from his role as Protector by the discovery that his wife, Eleanor, has been trapped into seeking occult help to divine the future. Arrested in 3.1, news of his murder emerges at 3.2.27 and his corpse is at once exhibited, bearing all the signs of a violent death (3.2.147–78). The other main plot-line in 3.2 is the liaison between Queen Margaret and the Duke of Suffolk. The scene begins with them joining Gloucester's enemies – Cardinal Beauford, Bishop of Winchester, and Richard Plantagenet, Duke of York – in trying to turn the King against the Protector. Henry affirms his trust in Gloucester, but when the news arrives of the loss of all the English territories in France Suffolk seizes the opportunity to arrest Gloucester on the charge of high treason. The King laments his downfall, powerless to prevent it, and once he has left the scene the Queen, Suffolk, and York endorse Beauford's offer to 'provide his executioner' (3.1.276). Their success is short-lived, however, for the Commons are outraged at Humphrey's murder, and Warwick denounces Suffolk and Beauford as the instigators. Under pressure from the Commons, the King banishes Suffolk (3.2.279–88), who soon meets his death, the murderers leaving 'his head and liveless body' for 'the Queen his mistress' to bury (4.1.142–3). Between Suffolk's banishment and death we witness Cardinal Beauford's guilt-ridden descent into madness and death (3.3).

The two remaining scenes with the highest density of unique Shakespearian matches are 4.2, which shows the rebel Jack Cade mustering his mob, and 5.1, where the Yorkist and Lancastrian forces confront each other before the battle of St Albans. The play ends in the middle of this battle, with two setbacks for the Lancastrian party: York kills Clifford, and his son Richard kills Somerset. The battle continues in *3 Henry VI*, which opens with the Yorkists counting their successes, Richard exhibiting Somerset's head.

The remarkable number of verbal matches between these two plays evidently derives from the shared subject matter, the disastrous War of the Roses, and the continuity of characters. As one family member dies, another succeeds him, and in the emblematic battle of Towton, we witness two deaths within a family, a father who has killed his son, and a son who has killed his father. A third reason for the linguistic homogeneity of the two plays is their closeness in time. As we move from the ending of one to the beginning of its successor, it seems as if Shakespeare had barely put down his pen. Having dramatized the whole sequence of war in *2 Henry VI*, from the opposed forces mustering for battle to its outcome for victors and losers, the phraseology of that play becomes a convenient template to describe each stage of the same process in its sequel. Here again, we see the preparations for battle from within an army, first marching and resting before it reaches its destination.⁶⁰ Henry VI gives way to Edward IV:

Well, **for this Night** we will repose us here: (2H6 2.1.196)

Now **for this Night**, let's harbor here in Yorke (3H6 4.7.79)

More often, the contending forces are described by those awaiting the enemy, or reinforcements:

The Duke of Yorke is newly come from Ireland,

And **with a puissant** and a mighty **power** ...

Is marching hitherward in proud array (2H6 4.9.24–7)

⁶⁰ In the quoted passages bold face indicates exact verbal parallels; underlined words and phrases fulfil the same syntactic or semantic function in both contexts.

The Queene is comming with a puissant Hoast, (3H6 2.1.207)

How farre off is our Brother Mountague?

By this at Daintry, **with a puissant troope** (3H6 5.1.5–6)

The Queene from France hath brought **a puissant power.** (3H6 5.2.31)

The linguistic templates of 2 *Henry VI* do duty in its successor, for the launching of troops towards the battlefield

They are all **in order, and march** toward us. (2H6 4.2.188)

Let's set our men **in order, | And issue** forth, (3H6 1.2.69–70)

and their imminent arrival:

Stafford and his brother **are hard by, with the Kings Forces.** (2H6 4.2.113–14)

She **is hard by, with** twentie thousand men: (3H6 1.2.51)

In both plays members of the warring factions affirm their loyalty, formally describing themselves in the third person:

War. **So God helpe** Warwicke, **as he loves** the Land, (2H6 1.1.205)

Mount. **So God helpe** Mountague, **as hee proves** true. (3H6 4.1.143)

And in both plays leaders appeal to their troops' loyalties:

And **you that** be the Kings Friends **follow me. Exit.** (2H6 4.2.181)

You that love me, and Warwicke, **follow me. Exit** (3H6 4.1.123)

Each side is conscious of their enemies' fortunes:

'Tis not enough **our foes are** this time fled, (2H6 5.3.21)

My Lord cheere up your spirits, **our foes are** nye, (3H6 2.2.56)

In the to-and-fro of combat, the participants are always dependent on news from other parts of the battlefield, and both plays have repeated requests for information and advice, in identical wording:

What counsaile give you in this weightie cause? (2H6 3.1.289)

What counsaile give you? whether shall we flye? (3H6 2.3.11)

What are they that thinke it? (2H6 3.1.107)

What are they that flye there? (3H6 4.3.28)

What newes? Why com'st thou in such haste? (2H6 4.4.26)

What newes? Why comm'st thou in such poste? (3H6 1.2.48)

How now? What newes? (2H6 4.4.26)

How now? what newes? (3H6 2.1.205)

The combat is not limited to England, for the territories in France are also at war, and news takes longer to arrive:

Welcome Lord Somerset: **What Newes from France?** (2H6 3.1.83)

Now Messenger, what Letters, or **what Newes From France?** (3H6 4.1.84–5)⁶¹

What answer makes your Grace to the Rebels Supplication? (2H6 4.4.7)

What answer makes King Lewis unto our Letters? (3H6 4.1.91)

The many matching collocations demonstrate the remarkable homogeneity of phraseology between *2 Henry VI*, and its successor. Shakespeare repeats the same phrases in order to chronicle not only the events of battle but the doings of individual characters. The largest group is clustered round the young King Henry, whose premature succession to the throne added to the political instability. In both plays Shakespeare makes Henry recapitulate this unfortunate event:

But **I was made a King, at nine months olde.** (2H6 4.9.4)

When **I was crown'd, I was but nine moneths old.** (3H6 1.1.112)

I was annointed King at nine monthes old, (3H6 3.1.76)

His youth necessitated the crucial appointment of a head of state, first Gloucester, then York:

Madame, I am **Protector of the Realme,** (2H6 1.3.120)

The Duke is made **Protector of the Realme,** (3H6 1.1.240)

As can be seen from those instances, Shakespeare gained additional emphasis by placing key terms at the end of a line, and using syntactic parallelism leading up to them. Beauford reminds his fellow plotters that, as Gloucester was the eldest living uncle to the then childless King,

he is the next of blood

⁶¹ See Alan Stewart,

And heyre apparent to **the English Crowne**: (2H6 1.1.151–2)

The armorer Thomas Horner is charged with the treasonous utterance that Richard Plantagenet, Duke of York,

Was rightfull Heire unto **the English Crowne**, (2H6 1.3.184)

At the play's beginning Warwick casts his power behind the Duke of York:

I'll plant Plantagenet, root him up who dares.

Resolve thee Richard, clayme **the English Crowne**. (3H6 1.1.48–9)

In the confusing oscillations of power in the later stages of the play, Warwick changes the destination of his patronage:

But Henry now shall weare **the English Crowne**, (3H6 4.3.48–9)

The source of the conflict propelling both parts of this play is the contested monarchy of Henry VI. Many parties desire him to abdicate, or else be deposed. The said Richard Plantagenet is determined to supplant him:

And force perforce Ile make him **yeeld the Crowne**, (2H6 1.1.258)

Warwick 'the Kingmaker', at this point supporting Plantagenet, asks Henry outright:

Wilt thou **yeeld the Crowne**? (3H6 2.2.101)

Plantagenet's eldest son, Edward, the leading Yorkist following his father's death, warns Henry of the dire consequences of his refusal to abdicate:

A thousand men have broke their Fasts to day,

That ne're shall dine, unlesse thou **yeeld the Crowne**. (3H6 2.2.127–8)

But the driving forces in both plays are determined to take what they want. First, the Duke of York resolves to seize the opportunity:

And when I spy advantage, claime **the Crowne**,

For that's the Golden marke I seeke to hit: 2H6 1.1.242–3

In the sequel his son, Richard Crook-back, makes the same resolve, but more cautiously:

And yet I know not how to get **the Crowne**,

For many Lives stand betweene me and home: (3H6 3.2.172–3)

Yet, a few lines later, he feels himself ‘like one lost in a thorny wood’, ‘toiling desperately’, observing himself

Torment my selfe, to catch **the English Crowne:** (3H6 3.2.179)

Shakespeare’s re-use of the identical syntactical structure suggests that in some cases he may have done so to create a deliberate parallel. Another echo that adds weight to this possibility is that between Suffolk’s indignant protest as he realizes that he is going to be killed by a mere lieutenant:

The honourable blood of Lancaster
Must not be shed by such a jaded Groome (2H6 4.1.51–2)

When Richard Crookback kills Henry VI he makes this sardonic comment:

What? will **the aspiring blood of Lancaster**
Sinke in the ground? I thought it would have mounted (3H6 5.6.61–2)

The name that occurs most frequently across both plays is that of the three successive Lancastrian Kings. York refers disparagingly to Bullingbroke,

Crown’d by **the Name of Henry** the fourth, (2H6 2.2.23)

When Clifford invokes that King’s illustrious son, conqueror of the French, the rebel Jack Cade foresees defeat as his followers change sides:

The name of Henry the fift, hales them to an hundred mischiefes,
and makes them leave mee desolate. (2H6 4.8.56–8)

Having first backed York, Warwick appeals to the soldiers to endorse the legitimate King against Edward IV:

Applaud **the Name of Henry**, with your Leader. (3H6 4.2.27)

By this accident of nomenclature, past Lancastrian Kings live on. Gloucester reminds the young King from whom he received the office of Protector:

As ere **thy Father Henry** made it mine; (2H6 2.3.34)

Yet Warwick can tell Edward, Prince of Wales, the son of Henry VI, that he has no right to the throne

Because **thy Father Henry** did usurpe (3H6 3.3.79)

Although his legitimacy is questioned, loyal subjects continue to support him:

This **is my King** Yorke, I do not mistake, (2H6 5.1.129)

And Henry **is my King**, Warwicke his Subject. (3H6 5.1.38)

Both of these affirmations, the first from Clifford, the second from Warwick, occur in the fifth act, as events move towards their – temporary – resolution.

In medieval political theory the King supposedly had ‘two bodies’: that of his office, which lived on in his successor, and that of his person.⁶² Shakespeare presents Henry as fully aware of the discrepancy between his office and his ability to perform it:

King. Was ever King that joy'd an earthly Throne,
And could **command no more** content then I? (2H6 4.9.1–2)

King. Let me entreat (for I **command no more**) (3H6 4.6.59)

In the first part of the sequence York declares his own superior fitness for the role:

I am far better born than is the King;
More like **a King, more** Kingly in my thoughts. (2H6 5.1.28–9)

Ironically, Henry uses the same form of words at the battle of Towton, where he has just witnessed two of the most savage events that civil war can produce, a son who has killed his father, and a father who has killed his son:

Sad-hearted men, much overgone with care
Heere sits **a King, more** wofull then you are. (3H6 2.5.123–4)

Shakespeare is unusual among dramatists of the 1590s in the care with which he described the symptoms of grief in response to suffering. So Queen Margaret claims to feel sorrow at the murder of Gloucester, to whom she was opposed:

And for myself, foe as he was to me,
Might liquid **teares**, or heart-offending **groanes**,

⁶² See E. Kantorowicz, *The King's Two Bodies*

Or blood-consuming sighes recall his Life;
I would **be blinde with weeping**, **sicke with grones**,
Looke pale as Prim-rose with blood-drinking sighes,
And all to have the Noble Duke alive. (2H6 3.59–64)

Shakespeare re-used several of these collocations, more economically, for the King witnessing the effects of war:

And let our hearts and eyes, like Civill Warre,
Be blinde with teares, and break **ore-charg'd with griefe** (3H6 2.5.78–9)

He deploys the same vocabulary of grief for King Henry, reacting to Cardinal Beauford's arrest of the good Duke Humphrey:

I Margaret: my **heart is drown'd** with **griefe**,
Whose **floud** begins to flowe within **mine eyes**; (2H6 3.1.198–9)

And he draws on this part of his phrasal lexicon again for Queen Margaret, describing to the French King Lewis her desperate situation since her husband is 'become a banished man':

From such a cause, as fills **mine eyes** with **teares**,
And stops my tongue, while **heart is drown'd** in **cares**. (3H6 3.3.13–14)

Both *2* and *3 Henry VI* are concerned with war and battles, but with a difference. The goal of many wars is territorial conquest, and the end product is victory. These two plays dramatize civil war, war within the same territory, where the ostensible goal is to depose the monarch from the House of Lancaster and replace him with one from the House of York. But as Shakespeare dramatizes the historical material, in the minds and motives of the participants, their primary goal that of killing their enemies. The word 'death' occurs with alarming frequency in these two plays and in their sequel. In *2 Henry VI* it occurs 68 times, in *3 Henry VI* it occurs 49 times, and in *Richard III* it occurs 67 times.⁶³ In many cases Shakespeare gives it greater emphasis by placing it at the line ending, often combined with

⁶³ These figures are taken from Marvin Spevack (ed.) *The Harvard Concordance to Shakespeare*.

related words. Cardinal Beauford, having had the good Duke Humphrey arrested, anticipates the next step to find a stratagem by which they can legitimize his murder:

But yet we want a Colour **for his death**: (2H6 3.1.236)

His fellow plotters dispute which of them has the greatest motive to wish the Duke dead.

Suffolk claims that ‘no man alive so fain as I’, but York disagrees:

’Tis Yorke that hath more reason **for his death**. (2H6 3.1.245)

In the fluctuations of power that mark the sequel York is captured by the Lancastrian forces, and in an extended sequence (3H6 1.4.66–180) Queen Margaret cruelly taunts him, offering him a napkin stained with the blood of his son, the young Earl Rutland:

And if thine eyes can water **for his death**,
I give thee this to dry thy cheeks withal (3H6 1.4.82–3)

In his defiant reply to this ‘She-wolf of France’ York breaks down weeping:

These tears are my sweet Rutland’s obsequies
And every drop cries vengeance **for his death**, (3H6 1.4.147–8)

The same line-ending can serve quite different functions. Clifford parades before the King Jack Cade’s captured supporters awaiting his verdict:

Expect your Highnesse doome of **life, or death**. (2H6 4.9.12)

In the sequel Shakespeare puts that absolute antithesis in the mouth of York, exulting over his sons’ prowess in battle:

they have demean’d themselves
Like men borne to Renowne, by **Life or Death**. (3H6 1.4.7–8)

From his reading of the chronicles Shakespeare realized the degree to which gratuitous cruelty emerges in civil war. Cardinal Beauford settles the dispute over who should kill the Protector:

But I would have him **dead, my Lord of Suffolke**, (2H6 3.1.273)

The sequel opens with the Yorkists triumphant, Richard Duke of Gloucester carrying the Duke of Somerset's head, which he proudly exhibits to his father. York uses the same phrasing to taunt the corpse:

But is your Grace **dead, my Lord of Somerset?** (3H6 1.1.18)

When Shakespeare re-uses a phrase from *2 Henry VI* in the sequel it is often, as here, applied to a much nastier purpose. In that play Warwick addresses the evil Cardinal, raving on his deathbed:

Beauford, it is thy Sovereigne **speakes to thee.** (2H6 3.3.7)

In its successor Warwick utters the same phrase, but he is now addressing the corpse of Clifford:

Speake Clifford, dost thou know who **speakes to thee?** (3H6 2.6.61)

Throughout both plays the human corpse is treated with callous disrespect. In this prolonged dance of death a rebel shares the same fate as a King. The worthy Alexander Iden of Kent, realizing that he has just vanquished Jack Cade, 'that monstrous traitor', despatches him to an inglorious fate:

Die damned wretch, the curse of her that made thee:
And as **I thrust thy body in** with my sword,
So wish I, I might thrust thy soule to hell.
Hence will I dragge thee headlong by the heeles
Unto a dunghill which shall be thy grave. (2H6 4.10.77–81)

Richard of Gloucester, having killed the King, disposes of the body:

He throw thy body in another roome,
And Triumph Henry, in thy day of Doome. (3H6 5.6.92–3)

In chronicling the rise of Richard, Duke of Gloucester in Part 3, Shakespeare twice repeats phrases he had used earlier for other characters but gives them added force in his mouth:

Hume. Hume must make merry with the Duchesse Gold:

Marry and shall (2H6 1.2.87–8)

Qu. Marg. Oh, kill me too. – Rich. **Marry, and shall.** (3H6 5.5.41–2)

He is prevented by King Edward – ‘hold, for we have done too much’. In *Part 2* Queen Margaret makes a grand gesture in her quarrel with the King over Humphrey’s death:

Qu. Marg. Aye me, I can no more: **Dye** [Margaret!]
For Henry weeps that thou dost live so long (2H6 3.2.120–1)

In *Part 3* Shakespeare uses the same syntax and phrasing for Richard, cutting off King Henry’s diatribe in mid flow:

He heare no more: | **Dye** Prophet in thy speech (3H6 5.6.57)

In this chronicle of many deaths, those observing a dying man dispassionately note his last movements (for the benefit of an audience that cannot see them). Warwick observes Cardinal Beauford’s death pangs:

See how **the pangs of death** do make him grin (2H6 3.3.24)

In the sequel Richard Duke of Gloucester uses the same phrasing to describe the death of Warwick’s illegitimate half-brother:

And in **the very pangs of death**, he cryde (3H6 2.3.17)

In the earlier play Suffolk, going to his exile and death, makes a last request to Queen Margaret:

To have thee **with thy lippes** to stop my mouth:
So should'st thou eyther turne **my flying soule**,
Or I should breathe it so into thy body (2H6 3.2.396–8)

In the sequel Shakespeare uses the same form of words to represent Warwick’s pathetic dying appeal to Montague – who is lying dead somewhere else on the battlefield:

If thou be there, sweet Brother, take my Hand,
And **with thy Lippes** keepe in my Soule a while. (3H6 5.1.35–6)

Since Shakespeare wrote in a period where Christianity was unquestioned, his characters have due awareness of the soul as a cognitive agent. Warwick tries to convince Henry that Gloucester was murdered:

As surely as **my soule intends** to live (2H6 3.2.153)

Shakespeare places the phrase in the mouth of King Edward, wooing Lady Jane Grey:

I speake no more then what **my Soule intends**, (3H6 3.2.94)

Gloucester swears his loyalty by the highest authority:

But God in mercie so deale **with my Soule**, (2H6 1.3.157–8)

Clifford, fatally wounded, expresses his loyalty to Henry for the last time.

As I in dutie love my King and Countrey.

O Lancaster! I feare thy overthrow,

More then my Bodies parting **with my Soule**: (3H6 2.6.3–5)

The pious King invokes God's benevolence over Beauford as he raves in delirium:

Peace to his soule, **if Gods good pleasure be**. (2H6 3.3.26)

He repeats that piety as he observes the terrible battle of Towton:

Would I were dead, **if Gods good will were so** (3H6 2.5.19)

The fact that in the Folio text characters in both plays invoke God so openly is a sign that they were written before the Act of 1606 making it an offence to do so in stage plays. Indeed, both plays record instances of oaths using the divine name, which were then wholly censored for public performance. Gloucester quarrels with Beauford:

Now **by Gods Mother** Priest,

I'll shave your crown for this, (2H6 2.1.49–50)

King Edward, wooing Lady Grey, again swears by the ultimate power.

And **by Gods Mother**, I being but a Batchelor, (3H6 3.2.103)

In both of these civil war plays Shakespeare uses animal imagery to dramatize the process by

which his protagonists seek power or revenge.⁶⁴ York, seeking out Old Clifford, whom he is about to kill, warns off a rival:

Hold Warwick: seek thee **out some other chace**

For I my selfe must hunt this Deere **to death.** (2H6 5.2.14–15)

I mentioned earlier that some of the verbal matches between the two plays are so close as to suggest that Shakespeare may have consciously echoed phrases that he had used before, either for the same character, or the same situation. This is surely the explanation for this remarkably close parallel, spoken by Richard Gloucester, York's son, as he pursues Clifford's son:

Nay Warwicke, single out some other Chace,

For I my selfe will hunt this Wolfe **to death.** (3H6 2.4.12–13)

In both plays Shakespeare makes use of the fact that the heraldic badge of Salisbury and his son Warwick was a rampant bear chained to a staff. So York summons their aid:

Call hither to the stake my **two brave Beares**,

That with the very shaking of **their Chaines**,

They may astonish these fell-lurking curs. (2H6 5.1.144–6)

That metaphor is drawn from a much less glamorous source, the bear-baiting arenas on the South Bank. Shakespeare repeats the connection again in the closing scene, but from a different perspective. King Edward, 'once more' restored to the throne, lists all the 'valiant foemen' whom he has 'mow'd down': three Dukes of Somerset, two Cliffords, two Northumberlands,

With them, the **two brave Beares**, Warwick & Montague,

That in **their Chaines** fetter'd the Kingly Lyon, (3H6 5.7.10–11)

Classical rhetoric taught that metaphors should be used for a specific purpose, often codified in the traditional categories of praise or blame. Queen Margaret one of the Lord Protector's

⁶⁴ See Audrey Yoder, *Animal Analogy in Shakespeare's Character-Portrayal* (New York, 1947).

enemies, invokes the lion and its invincible power to warn the King against the supposed threat that Gloucester represents:

Small Curses are not regarded when they grynne,
But great men tremble **when the Lyon rores**,
And Humphrey is no little man in England. (2H6 3.1.18–20)

In the sequel Henry dismisses Exeter's fear that the approaching army, raised by Edward, will attract more followers by appealing to the King's royal grace and recalling that idyllic biblical image of predators living in peace with their prey:⁶⁵

No Exeter, these graces challenge grace
And **when the Lyon fawnes** upon the Lambe,
The Lambe will never cease to follow him (3H6 4.8.48–50)

A moment later Edward and his soldiers burst in, taking 'the shame-faced Henry' off to captivity, like a lamb to slaughter. All these animal analogies juxtapose the aggressor and the victim. An added degree of vulnerability comes when hungry birds of prey are involved.

York uses this analogy to argue against the appointment of a rival to a position of power and influence:

Wer't not all one, **an emptie Eagle** were set,
To guard the Chicken from a hungry Kyte,
As place Duke Humphrey for the Kings Protector? (2H6 3.1.248–50)

In the sequel Henry, denounced by Margaret for having acceded to the Duke of York's demand that he should disinherit his son, hopes that Margaret can defeat York:

Revenge'd may she be on that hatefull Duke,
Whose haughtie spirit, winged with desire,
Will cost my Crowne, and like **an emptie Eagle**,
Tyre on the flesh of me, and of my Sonne. (3H6 1.1.266–9)

In this repeated analogy the associated phrases, 'hungry Kyte' and 'haughtie spirit', are

⁶⁵ C.f. Isaiah 11:16, 'The wolf also shall lie down with the lamb, and the leopard shall lie down with the kid; and the calf and lion together'. Of recent editors, Randall Martin is the only one to recognize this allusion: cf. *Henry VI, Part Three* (Oxford, 2001), p. 291.

linked by grammatical structure, not by meaning. In both plays the victim is notified about his predators, but also told that escape is impossible. The Duchess of Gloucester warns her husband that his enemies, Suffolk,

Yorke, and impious Beauford, that false Priest,
Have all lym'd Bushes to betray thy Wings,
And flye thou **how thou canst**, they'le tangle thee. (2H6 2.4.53–5)

The same warning is delivered with far greater force by King Edward to Warwick, kneeling in captivity:

Sayle **how thou canst**, | Have Winde and Tyde thy friend,
This Hand, fast wound about thy coale-black hayre,
Shall, whiles thy Head is warme, and new cut off,
Write in the dust this Sentence with thy blood,
Wind-changing Warwicke now can change no more. (3H6 5.1.53–8)

Many other passages in *3 Henry VI* reproduce the phrase structures of *2 Henry VI*, for a variety of purposes. In the earlier play Queen Margaret recalls her dangerous sea voyage to England for her arranged marriage, twice driven back by wind and waves:

And even with this, I lost faire Englands view,
And bid mine eyes be packing with **my Heart**,
And call'd them blinde and duskie Spectacles,
For losing ken of Albion's wished coast. (2H6 3.2.110–13)

When Shakespeare re-used this phrase structure in the sequel – three lines beginning with *And* (the rhetorical figure *anaphora*), the second line ending with the phrase 'my heart' – it could hardly have been in a more different context, Richard Crookback glorying in his ability to feign, or disguise emotion:

Why I can smile, and murther whiles I smile,
And cry, Content, to that which grieves **my Heart**,
And wet my Cheekes with artificiall Teares,
And frame my face to all occasions. (3H6 3.2.183–6)

That is one of the many instances of Shakespeare taking over a syntactical structure as a framework to be adapted for a new occasion. In both plays, characters invoke night as an extent of time to be whiled away. Suffolk, banished by the King, reveals his frustration to Queen Margaret:

Well could I curse away **a Winters night**,
Though standing naked on a Mountaine top, (2H6 3.2.335–6)

After the final and decisive defeat of the Lancastrian forces, Prince Edward, King Henry's son, exchanges insults with Richard by alluding to Aesop, who (like Gloucester), was deformed:

Let Aesop fable in **a Winters Night**,
His Currish Riddles sorts not with this place. (3H6 5.5.25–6)

Having also insulted Richard's brothers, King Edward and Clarence, all three stab the 'wilful boy' to death.

PROSODY

A further method by which we can contest Hugh Craig's claim that *2 Henry VI* was part-authored by Shakespeare and Marlowe is to examine its prosody. The growth of Elizabethan drama coincided with the rise of blank verse, introduced into English from the Italian by Surrey in his translation of the *Aeneid*, Books 2 and 4 (1540).⁶⁶ The iambic pentameter became the standard verse form for drama with the two great innovatory plays of 1587, Marlowe's *Tamburlaine* and Kyd's *The Spanish Tragedy*.⁶⁷ Those two pioneers differed in many respects, including the freedom with which they used blank verse. The strict pattern of alternating weak and strong stresses that make up the pentameter line was gradually loosened by adding at the end an extra unaccented syllable, the so-called 'feminine

⁶⁶ See, e.g., John Thompson, *The Founding of English Metre* (London, 19)

⁶⁷ All dates are taken from Martin Wiggins, *British Drama 1522–1642: A Catalogue*, 10 vols. (Oxford, 2012–).

ending' (a term taken from the French). In his pioneering study Philip Timberlake showed that Marlowe made little use of this resource, the total occurrence of feminine endings ranging from 1.3 per cent (*Tamburlaine*) to 1.7 per cent (*The Jew of Malta*) and 3.7 per cent (*Edward II*).⁶⁸ While Kyd's *Spanish Tragedy* showed the same strict style as *Tamburlaine*, Kyd soon developed greater freedom: *Soliman and Perseda* (1588) had 10.2 per cent, *King Leir* (1589) had 10.8 per cent and *Arden of Faversham* (1590) scored 6.2 per cent.⁶⁹ Timberlake commented that Kyd 'freely admitted feminine endings because he saw their fitness for dramatic speech' (53): Shakespeare followed his example. I reproduce Timberlake's analysis of feminine endings in *2 Henry VI* (88), which amount to 10.4 per cent, on his 'strict' criteria – that is, excluding 'proper nouns constituting feminine endings' (5). I have marked with an asterisk those scenes that Hugh Craig would remove from Shakespeare. It will be instantly obvious that they have the same proportion of feminine endings as other scenes of comparable length. Moreover, since the mean total for Marlowe never exceeds 3.7 per cent, it is evident that he had no hand in this play.

⁶⁸ See Philip W. Timberlake, *The Feminine Ending in English Blank Verse. A Study of its Use by Early Writers in the Measure and its Development in the Drama up to the Year 1595* (Menasha, WI, 1931), pp. 36–45.

⁶⁹ *Ibid.*, 46–53, 61–2. For new evidence on the Kyd canon see Brian Vickers, 'Thomas Kyd, secret sharer', *TLS*

2 Henry VI

<i>Act, Scene</i>		<i>Full lines</i>	<i>All fem. endings</i>	<i>Proper names</i>	<i>Fem. End., Strict ct.</i>	<i>Total per cent</i>	<i>Strict per cent</i>
*1.1		235	23	8	15	9.7	6.3
*1.2		103	15	4	9	14.5	8.7
1.3		164	17	3	14	10	8.5
1.4		50	6	4	1	12	2
2.1		137	15	2	10	10.9	7.3
2.2		79	12	6	6	15.1	7.5
*2.3		55	8	0	8	14.5	14.5
2.4		106	13	0	12	12.2	11.3
3.1		371	54	8	45	14.5	12.1
3.2		400	53	6	46	13.2	11.5
3.3		33	7	0	6	21.2	18.1
4.1		144	14	1	13	9.7	9
*4.2		50	1	0	1	2	2
*4.3		Prose					
*4.4		53	3	0	3	5.6	5.6
*4.5		7	1	0	0	14.2	0
*4.6		Prose					
*4.7		38	10	0	8	26.3	21
*4.8		34	7	0	7	20.5	20.5
*4.9		46	13	1	10	28.4	21.7
*4.10		40	10	0	10	25	25
5.1		208	26	1	22	12.5	10.5
*5.2		76	9	0	8	11.8	10.5
5.3		31	6	2	4	19.3	12.9
	Total	2,460	323	46	258	13.1	10.4

A generation later, the Estonian scholar Ants Oras studied the prosody of Elizabethan and Jacobean drama from a different aspect.⁷⁰ Oras studied the placing of pauses within the blank verse line as indicated by punctuation marks. As he explained,

The question I ask – too seldom asked by students of prosody – is not how often on the whole pauses occur in a work, but in what positions they appear in the verse, and in what ratios compared with other positions in the line. So, no attempt has been made to compute the number of pauses in relation to the total number of lines in a work, which has been the usual procedure. What has been studied is the incidence of internal pauses in each of the nine positions possible within an iambic pentameter line in relation to the totals of such pauses, regardless of the amounts represented by such totals. Plays with only a hundred internal pauses indicated by punctuation may thus present percentage patterns identical with those of the plays – or poems – having thousands of such pauses. Authors may deliberately choose to use little or much pausation in their verse, but they will generally be less aware of the positions in the line in which they pause. It is these less conscious pause patterns that I have attempted to find and analyse. Although it will become apparent that even in pause patterning there can be much deliberation, and even rigid regulation, especially as regards the main pause in a verse, the ‘caesura’, the total patterns are likely to reveal much over which the person concerned has little or no control, almost as people are unable to control their cardiograms. (1–2)

The method Oras used had to be statistical: ‘Only by meticulous counting and computation could the basic patterns be detected.’ Oras provided the raw data for ‘internal’ pauses within the line at each of the nine pausal positions, together with the total of internal pauses for each work.

The individual stress profiles are easier to grasp in the over seven hundred graphs that Oras provides, which

simply present the statistical data in a visual form. They are all drawn to the same scale. Height represents the percentage figures, whereas horizontal interval indicates the position in the five-stress, ten-syllable verse line – the first position, the second, the third, and so on through the ninth. I give here three examples from the A-sequence for Shakespeare, for *Titus Andronicus*, *Hamlet*, and *Cymbeline*. The nine percentages are:

1	2	3	4	5	6	7	8	9
<i>Titus Andronicus</i> : 4.9, 11.6, 5.4, 35.4, 17.1, 19.3, 4.7, 1.5, 0.1								

⁷⁰ Ants Oras, *Pause Patterns in Elizabethan and Jacobean Drama. An Experiment in Prosody* (Gainesville, FL., 1960). Oras subsequently published *Blank Verse and Chronology in Milton* (Gainesville, FL., 1966).

Hamlet: 2.5, 7.1, 3.8, 27.1, 14.9, 27.7, 11.4, 4.2, 1.1
Cymbeline: 2.0, 5.1, 3.3, 15.8, 12.5, 25.6, 15.2, 14.3, 6 .2.

Disregarding any decimals of fewer than five points and counting any five decimal points as a full unit, we obtain the three graphs shown here. The graph for *Titus Andronicus*, accordingly, indicates that approximately 5 per cent of all punctuation marks *within* the five-stress lines of the play occur after the first syllable, 12 per cent occur after the second syllable, 5 per cent after the third syllable, and none after the ninth syllable; this graph has nothing to say about punctuation at the end of the line. The figures to the right, close to the top of each graph, indicate the total number of internal pauses in each work. (5–6)

I quote this virtually unknown work at some length, because it illuminates a fundamental element in early modern verse drama, giving an immediately understandable record of each play's individuality. The graphs that follow will show that Shakespeare's *2 Henry VI* and *3 Henry VI* are homogeneous in their prosody, and very similar to *Richard III*. Taken together with Timberlake's analysis of feminine endings, we can be confident that Shakespeare wrote all three plays, and that Marlowe had no hand in them.

[For Oras's graphs see the attached Appendix (pdf)]

Bibliography

- Bengt Altenberg, 'On the phraseology of spoken English: the evidence of recurrent word-combinations', in A.P. Cowie (ed.) *On Phraseology. Theory, Analysis, and Applications* (Oxford, 1998), pp. 101–22.
- Charles Bally and Albert Sechehaye, *Cours de linguistique générale* (Saussure) (Paris, 1916).
- David Bevington and Eric Rasmussen (eds.), *Doctor Faustus. A- and B-Texts (1604, 1616)* (Manchester and New York, 1993) esp. pp. 62–73.
- D. Biber, S. Conrad and R. Reppen, *Corpus Linguistics: Investigating Language Structure and Use* (Cambridge, 1998).
- J.P. Brockbank, 'The frame of disorder – *Henry VI*', in J.R. Brown and Bernard Harris (eds.) *Early Shakespeare* (London, 1901), pp. 73–99.
- Geoffrey Bullough (ed.), *Narrative and Dramatic Sources of Shakespeare, Volume Three, Earlier English History Plays: Henry VI, Richard III, Richard II* (London and New York, 1960), pp. 95–8, 113–18, 128–33.
- John Burrows, *Computation into Criticism: A Study of Jane Austen's Novels and an Experiment in Method* (Oxford, 1987)
- '“An Ocean Where Each Kind...”: statistical analysis and some major determinants of literary style', *Computers and the Humanities*, 23 (1989): 309–21.
- '“Delta”: a measure of stylistic difference and a guide to likely authorship', *Literary and Linguistic Computing*, 17 (2002): 267–87.
- 'Questions of authorship attribution and beyond', *Computers and the Humanities*, 37 (2003): 3–32.
- 'All the way through: testing for authorship in different frequency strata', *Literary and Linguistic Computing*, 22 (2007): 27–47.
- F.M. Cornford, (trans.) *Sophist*, 261e–262d, repr. in *Plato. The Collected Dialogues*, ed. Edith Hamilton and Huntington Cairns (New York, 1963).
- Hugh Craig and Arthur F. Kinney, *Shakespeare, Computers, and the Mystery of Authorship* (Cambridge, 2009)
- 'A collaboration about a collaboration: the authorship of *King Henry VI, Part Three*', in Marilyn Deegan and Willard McCarty (eds.), *Collaborative Research in the Digital Humanities* (Farnham and Burlington VT, 2012), pp. 28–65.
- I. M. Crombie, *An Examination of Plato's Doctrines. Volume II. Plato on Knowledge and Reality* (London, 1963).
- John Dover Wilson, *The First Part of King Henry VI* (Cambridge, 1952).
- Alvar Ellegård, *A Statistical Method for Determining Authorship: the Junius Letters, 1769–1772* (Göteborg, 1962).

- Nick C. Ellis, 'Memory for language', in P. Robinson (ed.), *Cognition and second language instruction* (Cambridge, 2001), pp. 33–8.
- Robert Fabyan, *The New Chronicles of England and of France*, ed. H. Ellis (London, 1811).
- Darren Freebury-Jones, 'Exploring co-authorship in *2 Henry VI*', *Journal of Early Modern Studies*, 5 (2016): 201–16.
- Zellig S. Harris, 'Distributional structure', *Word*, 10 (1954):146–62.
- David I. Holmes, 'Authorship attribution', *Computers and the Humanities*, 28 (1994): 87–106.
- D.L. Hoover, 'Statistical stylistics and authorship attribution: an empirical investigation', *LLC*, 16 (2001): 421–44
- 'Frequent word sequences and statistical stylistics', *id.*, 17 (2002): 157–80
- 'Frequent collocations and authorial style', *id.*, 18 (2003): 261–86
- 'Testing Burrows's Delta', *id.*, 19 (2004): 453–75.
- 'Corpus stylistics, stylometry, and the styles of Henry James', *Style*, 41 (2007): 174–2003.
- G.K. Hunter *All's Well that Ends Well*, Arden edn. (London, 1959).
- E. Kantorowicz, *The King's Two Bodies*
- Patrick Juola, 'Authorship attribution', *Foundations and Trends in Information Retrieval* 1 (2006): 234–334.
- Deborah Levine Gera, *Ancient Greek Ideas on Speech, Language, and Civilization* (Oxford, 2003).
- John Lyons, *Introduction to Theoretical Linguistics* (Cambridge, 1971).
- Randall Martin, *Henry VI, Part Three* (Oxford, 2001).
- William Montgomery (ed.) *The First Part of the Contention 1594* (Oxford, 1985; Malone Society Reprints).
- Frederick Mosteller and David Wallace, *Inference and Disputed Authorship. 'The Federalist'* (Reading, MA, 1964). Introductory essay by John Nerbonne, 'The exact analysis of text' (Stanford, CA, 2001), pp. xi–xx.
- Martin Mueller, *The Iliad* (2nd edition; London,).
- J.R. Nattinger, 'Some current trends in vocabulary teaching', in R. Carter and M.J. McCarthy (eds.) *Vocabulary and Language Teaching* (London, 1988), pp. 62–82.
- Ants Oras, *Pause Patterns in Elizabethan and Jacobean Drama. An Experiment in Prosody* (Gainesville, FL., 1960).

- Randolph Quirk, Sidney Greenbaum, Geoffrey Leech, Jan Svartvik, *A Grammar of Contemporary English* (London, 1972), p. 305
- Joseph Rudman, 'The state of authorship attribution studies: some problems and solutions', *Computers and the Humanities*, 31 (1988): 351–65
— 'The state of non-traditional authorship attribution studies – 2012: some problems and solutions', *English Studies*, 93 (2012): 259–74.
- Ferdinand de Saussure *Cours de linguistique générale*, in Tullio de Mauro, 'Edition critique' (Paris, 1972, 1985), and in Roy Harris, trans. *Course in General Linguistics* (London, 1983).
- Marvin Spevack (ed.) *The Harvard Concordance to Shakespeare*.
- Efstathios Stamatatos, 'A survey of modern authorship attribution methods', *Journal of the American Society for Information Science and Technology*, 60 (2009): 538–56 (539).
- Alan Stewart,
- John Thompson, *The Founding of English Metre* (London, 19).
- Philip W. Timberlake, *The Feminine Ending in English Blank Verse. A Study of its Use by Early Writers in the Measure and its Development in the Drama up to the Year 1595* (Menasha, WI, 1931).
- Brian Vickers, *The Artistry of Shakespeare's Prose* (London, 1968; 2nd edn., London, 2005).
— *Shakespeare, Co-Author. A Historical Study of Five Collaborative Plays* (Oxford, 2002), 47–57.
— 'Incomplete Shakespeare: or, denying co-authorship in *1 Henry VI*', *Shakespeare Quarterly*, 58 (2007): 311–52.
— 'Identifying Shakespeare's additions to *The Spanish Tragedy*' (1602): a new(er) approach', *Shakespeare* 8 (2012): 13–43.
'Thomas Kyd, Secret sharer', *TLS*
'Repetition and emphasis in rhetoric theory and practice', in Andreas Fischer (ed.), *Repetition* (Tübingen, 1994), 85–113, with examples from *3 Henry VI*, Dickens, *Bleak House* (1853), and Beckett, *The Unnameable* (English version, 1958).
.....
— 'The two authors of *Edward III*', *Shakespeare Survey* 67 (2014): 69–84.
— 'Shakespeare and the 1602 Additions to *The Spanish Tragedy*: a method vindicated', id. 12 (2016).
— 'Giving Greene his *Groatsworth*', *Papers of the Bibliographical Society of America* (forthcoming).

— (ed.) *The Collected Works of John Ford*, vol. 2 (Oxford, in the press), Introduction: 'Dramatists repeat themselves'.

P.J. Vincent, 'Unsolved mysteries in *Henry VI, Part Two*', *Notes and Queries*, 48 (2002): 270–4.

Martin Wiggins, *British Drama 1522–1642: A Catalogue*, 10 vols. (Oxford, 2012–).

John Dover Wilson, *The First Part of King Henry VI* (Cambridge, 1952).

Audrey Yoder, *Animal Analogy in Shakespeare's Character-Portrayal* (New York, 1947).