Introduction
Gabriel Egan

In Liz Duffy Adams's play, Christopher Marlowe is William Shakespeare's co-author on *Henry VI Parts One, Two* and *Three*. Is there any truth in this?

Adams's play would be a remarkable achievement even if it were mere invention. But it has the additional merit of creatively exploring the implications of a newly discovered historical fact: that Marlowe created a little more than we used to think (and Shakespeare a little less) of the astonishing body of British drama from the late sixteenth century.

In a courtroom a witness is typically told to describe what happened, but 'in your own words'. The comedian George Carlin saw a problem here and asked his audience: 'Do *you* have your own words? I'm using the ones everyone else has been using.' It is commonly claimed that Shakespeare had his own words, in the sense that he invented new words that have entered our shared language. Recently it has been discovered that this is untrue.

We can search in digital databases of all British books published before and during his lifetime for each word that we suspect Shakespeare coined. Except for trivial exceptions, we find in every case that someone else used the word before him. The exceptions are such things as putting the prefix *un-* before an existing verb or noun. Most of Shakespeare's inventions of that kind did not catch on, including 'I am unkinged' (*Richard II*, Act 5, Scene 5), 'unsex me here' (*Macbeth*, Act One, Scene Five), and 'unshout the noise' (*Coriolanus*, Act 5, Scene 5).

Although writers occasionally coin new words, that is not what we usually mean by a writer's personal style. Language is fundamentally dependent on agreement about what the words mean, which limits the opportunities to coin wholly new ones.

Authorial distinctiveness, what we mean by style, lies not in the invention of individual words but in the inventiveness of the order of words chosen by an author.

It is in the nature of language that a writer need put only a few words together to produce a sequence of them that no one has used before. According to Google's search engine, the six-word sequence that I just used, 'the order of words chosen by', has appeared many times in the hundreds of millions of webpages it has indexed. But if we quote just one more word of mine, to make it 'the order of words chosen by an', Google reports never having seen this sequence before.

Publication of this essay might change that result, so the reader is encouraged to take a sample of her own writing and see how few of her words (presented within quotation marks) are needed to get a Google's search to report that it has not seen the sequence before. So few are needed that mere uniqueness of phrasing cannot be the essence of authorial style either. Yet it seems intuitive that authorial style must have something to do with how writers follow one word with another in sequences.

In *Born With Teeth*, Shakespeare confesses himself dazzled by Marlowe's way of combining words. Both were born in 1564, but Marlowe had the creative headstart and by his mid-twenties had written the hit plays *Tamburlaine the Great Parts 1* and *2*, *The Jew of Malta* and *Doctor Faustus*. Marlowe's writing was distinctive, widely admired, and in some ways quite easy to parody. Shakespeare parodied it in *Henry IV Part Two* when Ensign Pistol rants about 'hollow pampered jades of Asia' (Act 2, Scene 4), echoing 'Holla, ye pampered jades of Asia' from Marlowe's *Tamburlaine the Great Part 2* (Act 4, Scene 3). Enthusiastic theatregoers of Shakespeare's time, of which there were many, were probably expected to spot this recycled Marlovianism.

But imitation can also be homage rather than parody. Looking up at a beautiful young woman framed in a window, Barabas in Marlowe's *The Jew of Malta* says, 'But stay: what star shines yonder in the east? / The loadstar of my life, if Abigail' (Act 2, Scene 1). The same likening of a beautiful young woman in

a window to a brightly burning star occurs to Shakespeare's Romeo: 'But soft! What light through yonder window breaks? / It is the East, and Juliet is the sun!' (Act 2, Scene 1).

More subtle signs of Marlowe's influence recur across early modern drama, in the order of selected words, the use of rhythmic language, and in the memorable actions of characters. An example of the last of these is that from her window Abigail throws valuables down to Barabas in *The Jew of Malta*, just as Jessica at her window throws valuables down to Lorenzo in Shakespeare's *The Merchant of Venice*, the title of which also echoes Marlowe's title.

With Marlowe such an influential figure across early modern drama, admired and imitated by many, it is fair to ask how scholars can hope to distinguish Marlowe's style from that of the writers influenced by him, including Shakespeare. To understand that, the creation of the three *Henry VI* plays that Liz Duffy Adams thrillingly dramatises presents a central test case. In 2016, the General Editors of the New Oxford Shakespeare Complete Works edition from Oxford University Press put Marlowe's name alongside Shakespeare's on the title pages of *Henry VI Parts One*, *Two* and *Three*. How could we tell that Marlowe had contributed to these plays?

There is no external evidence that Marlowe had a hand in any of them. *Henry VI Part One* was first published in 1623, seven years after Shakespeare's death, in the first collected-plays edition of Shakespeare, entitled *Comedies, Histories, & Tragedies* and now commonly referred to as the First Folio. There were three more editions of this collection later that century, each reprinting its predecessor.

The First Folio collection of 1623 also included *Henry VI Part Two* and *Henry VI Part Three*. But almost thirty years earlier two plays had been published that are strikingly similar to them: one was called *The Contention of York and Lancaster*, published in 1594 and remarkably similar to *Henry VI Part Two*, and the other was called *Richard Duke of York*, published in 1595, which resembles *Henry VI Part Three*.

Scholars disagree about the relationships between *The Contention of York and Lancaster* and *Henry VI Part Two* and between *Richard Duke of York* and *Henry VI Part Three*. Over the centuries the theories offered have included that the plays published in 1594 and 1595 are early versions of two plays that Shakespeare went on to revise to make his *Henry VI Part Two* and *Henry VI Part Three*. Perhaps there was also an early *Henry VI Part One* and it is simply lost.

Another theory is that the plays we find in the 1623 Folio were already written and being performed in the early 1590s, but somehow they got mangled in the process of being printed in 1594 and 1595. Perhaps this happened because the publisher did not have authorised manuscripts of the plays, but only, as it were, pirated copies of the scripts made by some surreptitious process.

Both theories might be true at once. The two play editions of the 1590s might be corrupted versions of two early Shakespeare plays that also differ from their counterparts of 1623 because of authorial or non-authorial revision. Anyone who presents a definitive narrative about the relationships between the two play editions of the 1590s and the ones in the 1623 Folio is claiming more than we currently know (as of mid-2025) about the matter, although of course new information settling the question could emerge at any time.

So, who wrote these four plays? The 1623 Folio attributes to Shakespeare all thirty-six of the plays it collects: his name and the famous engraving of his face are prominent at the beginning of the book. But we know this is misleading, since everyone agrees that the Folio's *Titus Andronicus* is partly by George Peele, its *Timon of Athens* partly by Thomas Middleton, and its *Henry VIII* partly by John Fletcher. Two further plays, *Pericles* and *The Two Noble Kinsmen*, were omitted from the 1623 Folio, but are widely accepted as Shakespeare's collaborations with George Wilkins and John Fletcher respectively.

No author is named on the title pages of *The Contention of York and Lancaster* (1594) and *Richard Duke of York* (1595). This anonymity surprises modern readers, but was not unusual

at the time. For a modern analogy, consider that writers of today's Hollywood screenplays often will not get a prominent credit for their work. We usually discuss the film *Taxi Driver* as Martin Scorsese's achievement as its director rather than Paul Schrader's, as its screenwriter.

Shakespeare's name did not appear on the title pages of any of his published plays until 1598, but this should arouse no suspicion. At the start of Shakespeare career around 1588 it was simply usual for a printed play's title page to omit to identify its author, and by the end of his career around 1613 it had become usual to mention the author.

For some works published anonymously we have other external evidence of who wrote what. In 1598, Francis Meres's book *Palladis Tamia* helpfully listed Shakespeare's hits up to that point, but it makes no mention of the *Henry VI* plays nor *The Taming of the Shrew*, which we think had been publicly performed by then. Sometimes official records for theatrical performances at court or as preparation for publication name a play's author(s).

When all such external evidence is absent, we fall back on the internal evidence of the text itself in order to attribute its authorship to one or more of the plausible candidates who were alive and in the right place at the time. The challenge is to find likenesses between the text we want to attribute and the writings of the possible candidates, and these must be likenesses that can plausibly be explained only by shared authorship and not by deliberate or unconscious imitation.

First we need a 'ground truth': a set of each candidate's works that we are confident are solo efforts. We look for the likenesses between these and the work we are trying to attribute. The best evidence is the preference shown by each author for overusing some words and phrases – and avoiding others. These words and phrases must not be too distinctive, since that makes them imitable and hence they will turn up in others' writings for that reason alone. An author beginning her novel with the words, 'It is a truth universally acknowledged…' is knowingly imitating the opening of Jane Austen's *Pride and Prejudice*.

Literary imitation might easily throw off our attempts to
identify authors by their preferred words and phrases. We would
not want to mistake Shakespeare's use of Tamburlaine's phrase
'pampered jades of Asia' or his borrowing of Barabas's starlight
metaphor for female beauty as evidence that Marlowe actually
wrote *Henry IV Part Two* or *Romeo and Juliet*. So how, then,
do we distinguish shared authorship from the conscious or
unconscious borrowings of one writer from another?

Sixty years ago, two statisticians made an important discovery
that improved the accuracy of authorship attribution. Working
on the problem of exactly who wrote each of the so-called
Federalist Papers – a series of essays by American Founding
Fathers Alexander Hamilton, James Madison and John Jay – the
statisticians Frederick Mosteller and David L. Wallace found
that writers most strongly reveal their style in their preferences
for the least interesting words.

About one word in every eighteen words of the essay you are
reading is a '*the*' and about one in thirty-five is an '*and*'. The
top one hundred most-common words account for over half of
all that we say and write. With some minor variations, we all
use the same hundred common words so frequently that we
barely notice them. But we differ on how often we use each of
them, overusing some and underusing others.

Nobody seems to be conscious of these preferences, but they
are demonstrably real and persistent – scarcely changing over a
lifetime – and they form an aspect of writerly style that appears
to be inimitable. If we have enough writing samples and a fairly
small list of candidate authors, authorship attribution questions
can be answered with about eighty to ninety per cent reliability
simply by counting the frequencies of these most-common words.

The New Oxford Shakespeare went a step further by also
tracking authors' habits in clustering these common words. We
recorded how often various authors put '*the*' near to '*and*', '*on*'
near to '*in*', and so on for every possible pair of the hundred
most-common words. We did this for all the words in all the
plays that we are sure are by Shakespeare and likewise in all
the plays by his contemporaries who have left us large enough
canons to test.

It is possible to measure how well these habits reveal authorship. The trick is to leave one play out of the process of generating the profiles of preferences for each candidate, and then to treat that play as if it were a play of unknown authorship that we want to attribute. We run this test many times, leaving out a different play each time, and we count how often the method points to the man we know actually did write the play. At best, our new method reaches about ninety-four per cent accuracy in its determinations, which is currently the state of the art.

No evidence exists to contradict Adams's conjecture that Shakespeare and Marlowe actively co-wrote the plays, with (as her play mentions) further contribution from Thomas Nashe. We were able to confirm what previous investigators had suggested: that Marlowe's writing style is present in *Henry VI Parts One, Two* and *Three*. How did it get there? We are unable to say. Textual scholars and historians of the theatre are free to generate theories of how the scripts got written, rewritten, passed between play companies, adapted, revived, stolen and published. We may never know exactly how Marlowe's writing got into these plays, but it is undoubtedly there.

Gabriel Egan is a General Editor of the New Oxford Shakespeare Complete Works edition, and at De Montfort University he teaches and researches on computational methods for better understanding the canon of English Literature.