

Non-Traditional Authorship Attribution Studies of William Shakespeare's Canon: Some Caveats

Joseph Rudman

Carnegie Mellon University (<jr20@andrew.cmu.edu>)

Abstract

The paper looks at the problems in conducting non-traditional authorship attribution studies on the canon of William Shakespeare. After a short introduction, the case is put forth that these studies are 'scientific' and must adhere to the tenets of the scientific method. By showing that a complete and valid experimental plan is necessary and pointing out the many and varied pitfalls (e.g., the text, the control groups, the treatment of errors), it becomes clear what a valid study of Shakespearean non-traditional authorship attribution demands. I then come to the conclusion that such a valid study is not attainable with the limits of present-day knowledge.

Keywords: Attribution, Authorship, Shakespeare, Statistics, Stylistics

It is not possible, in the compass of a single essay, to deal with very many – let alone all – of the tests by which investigators in their wisdom or folly have sought to prove authorship by style. (Schoenbaum 1966, 197)

1. *Introduction*

There are a few 'givens' framing this paper:

- 1) William Shakespeare was an actor and playwright – exactly who he was is not relevant here.
- 2) The First Folio constitutes the basis of what has come down to us as Shakespeare's canon.¹

¹ Non-traditional authorship attribution studies are those that make use of stylistics, statistics, and the computer. For a short history and overview of the field, see Rudman 1998, 2006, 2011, 2012. It is a tenet of non-traditional authorship attribution that the printed input texts (post-performance) used should be the closest to the author's final holograph (pre-performance) – ideally the practitioner should have both in hand. However, there is a large and unknown time gap between the final holograph of a play and that play printed in the First Folio, a potentially fatal flaw. Other plays such as *Pericles*

- 3) The non-dramatic works only will be dealt with peripherally.
- 4) No non-traditional authorship attribution study should be undertaken until an exhaustive traditional study is finished.
- 5) Non-traditional studies will return probabilities, not certainties.
- 6) Most of the points made below will not be explicated in depth (that awaits a forthcoming monograph).²
- 7) The bibliography ("Works Cited") is representative, not exhaustive.

This (obviously) is not the first paper to put forth caveats to authorship attribution on the Shakespeare canon. Samuel Schoenbaum gave his famous seven principles for the attribution of Elizabethan plays – and he said these do not exhaust the possibilities (1966, 191-197). Ward Elliott and Robert Valenza gave five 'Cautions and Caveats' (1996). Harold Love gave some caveats and an overview of the 'Shakespeare problem' in Chapter II of his book (2002a, 194-208). There are many other papers that survey the field and give caveats – none of these surveys are complete and no list of caveats is complete or even adequate. For example, Michael Oakes, in his book published in 2014, does not mention Thomas Horton, Sir Brian Vickers, or Marcus Dahl, among others, nor does he mention the pitfalls in textual selection (99-147). Gary Taylor (1987a, 1987b), Ian Lancashire (2002), Brian Vickers (2011), John Burrows (2012), MacDonald P. Jackson (2014), and Hugh Craig and Brett Hirsch (2014) – all leading practitioners in the field – have written surveys and critiques that are must-reads for anyone who wants to understand the field of non-traditional authorship attribution of Shakespeare's canon.

But Hope's methods are so flawed that all of his results are called into question ... Hope's insistence that his methods are 'more reliable' than 'other current approaches to authorship' (xv) is rather curious. (Rasmussen 1997, 111-112)

Although such teething troubles [problems with Morton's work] make it impossible to place any reliance upon current [1987] stylometric studies, they do not justify wholesale dismissal of the potential validity of such analysis. (Taylor 1987b, 80)

Almost all of the practitioners in the field are upbeat about the advances and successes in Shakespearean attribution studies. However, these studies are fraught with conflicts – conflicts over methodology and results. For example, Lancashire (2002) lists:

that did not appear in the First Folio but were added later to the canon are treated in my upcoming monograph.

² The monograph also contains in-depth critiques of most of the extant non-traditional authorship studies of the Shakespeare canon.

- 1) Merriam, Mathews, and Ledger vs. M.W.A. Smith
- 2) Foster and Abrams vs. Wells and Vickers
- 3) Charles Heiatt and Kent Heiatt vs. Foster

In addition there are:

- 1) Vickers (2011) vs. Craig and Kinney (2009)
- 2) Burrows (2012) vs. Vickers (2011)
- 3) Vickers and Dahl (2012) vs. Maguire and Smith (2012)
- 4) Rasmussen (1997) vs. Hope (1994)
- 5) Taylor (2016) vs. Stern (2004)
- 6) Craig vs. Vickers and Jackson (Craig and Hirsch 2014)
- 7) Taylor vs. Waugaman (Reisz 2014)

And this does not exhaust the controversies.

Gray Scott points out that Donald Foster takes Ward Elliott and Robert Valenza to task for inconsistent editing, failure to control for chronology, giving insufficient information for replicability, and miscounting (2006, paragraph 2). When readers who are outside of the community of practitioners read these back-and-forth articles (some of which cross over into ad hominem attacks), they get an overview of a discipline in disarray. What are non-experts expected to believe? Should they simply ignore everything and wait for some kind of consensus on methodology, or look at each study, each result, and form an educated opinion on the results? These are not intellectual lightweights trading criticisms that cast real doubt on each other's work – Vickers, Jackson, Burrows, and Craig, among many others, are at the top of their disciplines. They criticize each other's work but do not do so with enough detail to make the arguments clear and closing. There is not enough agreement on the basics behind the disputes, such as the final input texts and the statistical methodology, to convince each other, let alone the less sophisticated reader.

What follows throughout this paper hopefully will give readers some points of reference with which to judge the validity of the non-traditional authorship attribution studies of the Shakespeare canon.

2. Science and the Scientific Method

Uncertainty about whether stylometry is a science or not is further compounded by differences over what constitutes a science ... One view is that stylistics becomes scientific when it argues by means of numbers; but this is to take a very restricted view of science which leaves out experimental method, means of confirming hypotheses, repeatability of results, the capacity to induce universal laws from particular data, and, most importantly, the power to generate explanations. (Love 2002, 157)

The purpose of this dissertation is to explore the possibilities of using a scientific and statistical approach to solve a Shakespearean attribution question. (Horton 1987, 1)

The history of attribution studies ... showed that, when properly executed, studies using internal evidence were empirically sound, based on the careful collection and evaluation of data and using computer processes that could be replicated by other researchers – all attributes of a scientific methodology. (Vickers 2004, 106)

Despite Kinney's confidence in his method, computational stylistics once more gives the illusion of scientific procedure but yields no useful results. (Vickers 2011, 127)

If quantitative methodology can be shown to have improved, there are still humanities scholars who are wary of what they see as misplaced scientism in the importation of scientific methodology to literary studies. (Hope and Whitmore 2014, 2)

We may rightly be troubled when scientific ideas suffer uncertain translation into popular discourse, where long after being discarded by scientists, they seem to have an independent life with undeserved authority derived from their scientific origins. An example is provided by some work in computational stylistics, where antiquated notions of experimentation have hardened into doctrine (Rudman 1998; contra Burrows 2008). (McCarty 2011, 274)³

I would argue that attributional stylometry, as it now exists, is not a science in the sense claimed for it by a large party of its practitioners. (Love 2002, 161)

What stylometry offers, then, is not a science but a mathematisation of stylistics – a new way of discriminating between forms of language behaviour that is of great potential value but not as yet a way of accounting for them. (Love 2002, 160-161)

This paper is predicated on the premise that every non-traditional authorship attribution study is an experiment, a 'scientific' experiment. However, there is no universal agreement on which scientific principles (if any) should be invoked. The above quotations give a little glimpse of this. When practitioners use stylistics, statistics, and artificial intelligence techniques, they venture into the realm of science – albeit somewhat of a mutant science but nonetheless one demanding certain elements of the scientific method. Two of these elements are:

- 1) Reproducibility
- 2) An experimental plan

³ There was a pertinent and lively discussion of the topic of scientific method in literary studies started by Willard McCarty on the Humanist Discussion Group that began on 20 February 2015 and continued for some days. You can find this discussion at <dhhumanist.org>, accessed 28 February 2016.

2.1 *Reproducibility*

The joint use of these two procedures [giving the exact text that was used as input data and the computer programs used] fulfills one of the requirements of proper scientific method, namely the replicability of experiments. (Vickers 2011, 140)

Reproducibility is the backbone of any non-traditional authorship study. I have yet to see an argument against the concept. However, most practitioners do not give enough information so that another practitioner can re-run the experiment – exactly. In fact, I have not read one study of the Shakespeare canon that gives complete information, information that would allow me to replicate the study. Craig does this better than most in his publications and seems to become more complete in successive studies. But even Craig does not list the information for each play used in his studies; e.g., he says they are ‘from early printed editions’, that his electronic texts are from ‘online sources such as Literature on Line [LION] whenever possible,’ and that others were ‘keyboarded’ (Craig, Moscato and Rosso 2009, 918). On the other hand (except in the case of Donald Foster and his SHAXICON⁴), I have always been able to get unpublished details that would be needed for replication; e.g., Hope answered my email query about which Shakespeare texts he used.

This is not a concept that I alone have espoused over the years. Lancashire (2002) wrote about how one Shakespearean scholar should be able to reproduce the results of another scholar using the same style markers and same statistical tests.

There is another concept to be considered: duplicating an experiment. For this paper, *replication* means to follow the experimental plan of the original study in every detail without the slightest deviation; *duplication* means to reproduce the results using a different experimental plan, such as different style markers, different statistical tests, different control groups. Both are valuable, both are necessary. Only with replication and duplication of valid experiments that continually give the same results should a questioned play be admitted into the Shakespeare canon.

2.2 *Experimental Plan*⁵

Perform the following steps when designing an experiment:

- 1) Define the problem and the question to be addressed
- 2) Define the population of interest
- 3) Determine the need for sampling
- 4) Define the experimental design. (SAS 2005, 2)

⁴ SHAXICON is a lexical database that indexes all the words that appear in the [Shakespeare] canonical plays 12 times or less. <<http://goo.gl/qinxN9>>, accessed 28 February 2016.

⁵ See Rudman 1998 for a treatment of this concept that includes some references to experimental plans in linguistics.

An *experiment* is a process or study that results in the collection of data. (SAS 2005, 1; emphasis in original)

Experimental design is the process of planning a study to meet specific objectives. (SAS 2005, 1) (Emphasis in original)

The subsets of the experimental plan are:

- a) Input data (the texts)
- b) Controls
- c) Eliminate or control the variables not being tested
- d) Choice of style markers
- e) Other choices
- f) Statistical tests
- g) Sample selection and size
- h) Treatment of errors
- i) Analyze and interpret results

2.2.1 *The Texts*

No extant play excluded from the [First] Folio has ever been convincingly attributed in its entirety to Shakespeare; no play included in the collection has ever been convincingly attributed in its entirety to someone else. (Taylor 1987a, 36)

In order to carry out a valid non-traditional attribution study on the canon of William Shakespeare, you must have a body of ‘known’ Shakespearean texts and ‘known’ control texts, the latter being dependent on the particular study. What this means is that every word – every word – in the known Shakespeare be by Shakespeare! For the purposes of this paper, I will posit that the 1623 First Folio (third form)⁶ contains no non-Shakespearean play (collaborative and interpolative aspects of the First Folio will be treated later).⁷ One of the best discussions of text selection for Shakespeare studies (including non-Shakespearean control texts) is still Horton’s thesis (1987, 23-24).

A bug-free text of Shakespeare is a logical impossibility, since there is now general agreement that the texts are ineluctably multiple, and that in many cruces there can be no final ‘accurate’ version. (Best 2007, 155)

In their efforts to reproduce the words of a manuscript, compositors can commit any of the errors to which all copyists are liable: misreading, eye skip, dittography,

⁶ There is no *Troilus and Cressida* in the first form and no ‘Preface’ to *Troilus and Cressida* in the second form; see Blaney 1991, 14.

⁷ By using the First Folio, the practitioner can mitigate the need to unedit, de-edit, and edit. And, as will be shown later, there is a need to do parts of this process. For a more complete discussion, see Rudman 2005, 2012.

haplography, transposition, sophistication, substitution, simple omission, simple interpolation. But they add to these errors others specific to the medium of print. (Taylor 1987a, 43)

The concerns with the First Folio containing what Shakespeare actually wrote (known Shakespeare) do not end here. There is the matter of the transmission of his first manuscript through to the copy of the First Folio to the electronic (digitized) copy used by the attribution practitioner. For a good overview of most of the transmission processes and problems with ‘fowle papers’, ‘fayre copies’, and manuscripts in general, see Ioppolo (2006). Taylor gives a good introduction to the ‘problem’ of the Shakespearean texts and talks about the ‘permutations of dramatic manuscripts to print’ (1987a, 31). The following is a list of some of the transmission and ‘authorship’ problems that face the non-traditional attribution practitioner:

1) The many differences within the First Folio:

Charlton Hinman’s introduction to his 1963 *The Printing and Proof-Reading of the First Folio of Shakespeare* gives a good explanation of the many ways that the Shakespeare we have in the First Folio was changed from Shakespeare’s original manuscript. And it is not a consistent or constant change across all of the plays, in that every play [in the First Folio] presents its own unique problems’ (5).⁸ Hinman gives examples of press corrections made during the printing: ‘sining → sighing → singing’ (18); ‘botk → both, flelow → fellow’ (230); ‘evens → events, who → why, thy → try, namelesse → nameless, followes → follow, take → talk’ (253-254). Hinman states: ‘There are hundreds of variants ... such changes as it did produce tended rather to corrupt than to recover and preserve what Shakespeare wrote’ (I, 227). Horton, among others, points out that compositors varied spellings to justify lines.

Linguistic: Variant but different verbal forms: *ye* instead of *you*, *has* instead of *hath*, *between* instead of *betwixt*, contractions of pronouns and auxiliary verbs (*you’ll*, *I’m*), and so forth ... cannot be explained away as the result of deliberate – or even unconscious – imitation ... One weakness of such evidence is its occasional susceptibility to sophistication by certain scribes. (Taylor 1987b, 80)

Taylor states: ‘Shakespeare wrote, at a conservative estimate, at least 90 percent of the words included in the Folio’ (Taylor 1978b, 73). This appears to refer to the major collaborative parts and not word for word.

⁸ To make the problem even more complicated, the printing process on the First Folio was started ‘not later than August 1621 and was interrupted for more than a year’ (Hinman 1963, I, 16).

Oakes discusses the fact that there is not universal agreement that Shakespeare wrote all of the First Folio (2014, 100). It may seem nit-picking to worry about every word, but words, for the most part, make up the bulk of the style markers. Another caveat to practitioners is to make very clear what they mean by a ‘word’ – this is not the simple concept it might seem.⁹

2) The collaborative and interpolative parts in the First Folio:¹⁰

All plays ... are in a sense collaborations, shaped from conception to performance by the author’s awareness of the resources of actors and theatre, the wishes of the impresario or shareholders, and the tastes and capacities of the audience. (Schoenbaum 1966, 188)

There is no doubt that the principle of imitation took precedence over the idea of originality in early-modern compositional practice. (Kositsky and Stritmatter 2013, 141)

Furthermore, the collaborative project in the theatre was predicated on erasing the perception of any differences that might have existed, for whatever reasons. (Masten 1997, 17)

One message for attributionists is that any attempt to establish a database of assured ‘Shakespearean’ usage and parallels has to involve careful assessment of the originality of the passages tested. (Love 2002, 197)

Above all, this picture of the plays as Frankenstein’s monsters put together from different authorial parts denies the possibility of a Shakespearean voice, and in so doing robs the plays of what is, for most of us, their main interest. (Love 2002, 208)

Collaboration is ... a dispersal of author/ity, rather than a simple doubling of it; to revise the aphorism, two heads are different than one. (Masten 1997, 19)

Barry Clarke, in his Ph.D. dissertation (2013), makes a strong case that Francis Bacon ‘contributed’ small but significant interpolations to three of the plays in the First Folio: *The Comedy of Errors*, *Love’s Labour’s Lost*, and *The Tempest*. To have a valid study, this and all like interpolations must be removed from the input texts or a systematic error calculated; this calculation is almost impossible because we do not have sufficient examples.

⁹ For a discussion of this, see Rudman 2005. After finishing this paper it will become clear that studies of Shakespeare’s vocabulary are, for the most part, meaningless. Bradley Efron and Ronald Thisted’s well-known estimated number of words that Shakespeare knew but did not use (35,000) has no validity (Efron and Thisted 1976).

¹⁰ The treatment of collaboration in this paper is, by necessity, shortened. Vickers has done much excellent work on collaboration that is treated fully in the monograph; see Vickers 2002 and 2007, as well as Jackson 2003, another excellent work that will receive its due.

Rewriting and collaboration are much bigger problems, ones that may be insurmountable. Of the 36 First Folio plays, at least 16 have been cited by reputable scholars as being collaborations or at least having significant interpolations: 5 of 14 comedies, 7 of 10 histories, 4 of 12 tragedies. There are so many different kinds of collaboration that we will never be able to delete them out of the Shakespeare plays.¹¹

3) The changes made to the original text by actors, directors, scribes, and censors:

It should also be clear that critics need to account for the relentless change texts are subjected to as they pass through various theatrical and textual networks. (Farmer 2002, 173)

Posthumous adaptations may have occurred occasionally; censorship did occur systematically. (Taylor 1987a, 15)

It is crucial that we eliminate all of these changes before finalizing the input text. But we cannot. This is because we cannot identify the vast majority of them. We cannot subtract out this background noise. Therefore, we cannot know if we have a 'pure' sample – a potentially fatal flaw.

4) Quotations, languages other than English, and miscellany.

There is an overall substantial amount of text in the First Folio that is not in English. This should all be deleted before a non-traditional study is undertaken. For example, there are more than 200 words in French in *Henry V*, 3.4, and even more French interspersed throughout. There is also Latin in the plays. See for example *The Merry Wives of Windsor*, 4.1 where there is the *hic, haec, hoc, horum, harum, horum, qui, quae, quod* back and forth.¹² Any identifiable quotations should also be deleted, as well as anomalies such as Dr. Caius' dialogue in *The Merry Wives of Windsor*: 'By gar, de Herring is no dead so as I vill kill him' (2.3.12-13)¹³. This is a slippery slope and every decision to delete must be documented so the study can be replicated. The linguistic principles governing this kind of dialogue are quite different from standard English. If these Type Two style markers are not deleted (the total

¹¹ Example: Author A alternates the composition of scenes or acts with author B; Authors A and B sit and write together; Author A writes, using a rough draft from author B.

¹² Quotations, foreign languages, and the like are Type Two style markers and are not to be used in a statistical analysis of an author's style. See Rudman 2005 for a discussion of Type One and Type Two style markers.

¹³ Reference is from the Folger Library 1623 Folio.

number of words contained in these is substantial) or if a systematic error is not calculated, the attribution study will have a serious, if not fatal, flaw.

5) Other items to be removed.

Shakespeare, of course lifted plots and passages from Chapman's *Homer*, North's *Plutarch*, Golding's *Ovid*, Holinshed's *Chronicles*, and a host of other sources. (Groom 2003, 79)

Some sources have disappeared (e.g., the Jew-play mentioned by Stephen Gosson). (Bullough 1957, ix)

There is no doubt that Shakespeare used sources for his plays – many and varied. In his general conclusion, Geoffrey Bullough emphasizes the multiple possible sources for the various Shakespeare plays and stresses that we do not know which ones Shakespeare used (1975, 341-405). Kenneth Muir tells us that 'Shakespeare picked up "moldwarp", "dragon", and "lion" from Holinshed' (1978, 92). And a constituent problem is the question of at what point do translated words or paraphrases become borrowed words?

Bullough tells us that *The Famous Victories of Henry the Fifth* (1598 Anon.) was a source for Shakespeare's *1 and 2 Henry IV* and *Henry V* (1962, 155ff, 249ff, 347ff) and goes on to point out that the phrase, 'God knowes my sonne' is copied over by Shakespeare as, 'God knows, my son' (256, 318). Joseph Satin points out the famous line, 'a horse, a horse, a fresh horse' (1966, 2). The fact that Shakespeare uses sources does not, of course, denigrate his work; I have not seen one instance where the Shakespeare rework is not markedly better than the source. My point is that if you are doing a study of Shakespeare's style, these borrowings are Type Two style markers and should not be used. I realize this point is moot.

I have not seen a comprehensive study where a list of 'borrowed' words is produced. We do not know what effect the deletion of these words would have. There might not be a right or wrong way to treat this problem, but the practitioner must tell us exactly how each of the identifiable borrowed words and phrases are treated so the study can be replicated and critiqued.

Another item to be removed from the input text is the music. There is no real doubt that all of the music appearing in the First Folio should be eliminated before a stylistic analysis is undertaken. Not only is music a different genre, most of the music is either formulaic (e.g., fanfares) or of questionable authorship.¹⁴ Taylor questions, 'Did Shakespeare write the songs in *Measure for Measure* and *Macbeth*, or were they interpolated for a

¹⁴ I am not going to discuss the need to delete all of the stage directions and character names here. As far as I can tell almost all of the practitioners already do it.

posthumous revival?’ (1987b, 73). Noritaka Tomimura (2009) has a good analysis of the types of music in Shakespeare.

Surely the young Shakespeare did not start writing with the polish, expertise, and genius that is exhibited in his mature drama.¹⁵ Surely he had to pass through an apprentice phase where he learned to hone his craft. Surely he had to pass through a stage where he re-wrote and polished, where he collaborated as a junior partner. What is important is that we do not know the length, the years, or the output of these phases. If this were a court of law, the evidence (First Folio) would be inadmissible as the chain of custody was broken (shattered!) and the potential for contamination too great. Continuing in this vein and looking at the First Folio from a forensic linguist’s point of view, we must consider if the preponderance of evidence tells us that the First Folio is a pure enough sample of Shakespeare’s writing to allow a valid authorship study. I believe it is not.

Then, when you start narrowing down the number of plays because of genre and chronological constraints – and a further narrowing down because of the amount of text in each play to be eliminated due to collaboration, interpolation, and Type Two style markers – there is not enough text left to perform a valid study.¹⁶ However, the First Folio is the best we have. Each practitioner must decide whether or not it is sufficient. I think not, but feel that the vast majority (if not all) of the practitioners working with Shakespeare disagree. And remember that this text problem is only one of the caveats.

2.2.2 *Controls*

Since many plays were published anonymously, it is often exceptionally difficult to work out which playwright or playwrights wrote which script. (Bate 2013, 16)

There are various types of control that should be mentioned. The first is the concept of making sure it can be shown that there are no other playwrights in England at that same time (e.g., +/- five years) who have the same characteristics that were determined to be Shakespeare markers. Sara El Manar El Bouanani and Ismail Kassou (2014) add age, nationality, and gender as other factors to be controlled.

The way that this type of control should be carried out to obtain a valid non-Shakespeare control corpus and perform the necessary steps is: 1) hold out

¹⁵ This is a common theme in many articles and books on Shakespeare.

¹⁶ For example, if the unknown play being studied for authorship is a history play, we drop from ten plays to three simply by eliminating the seven ‘collaborations’ from the known Shakespeare.

a random sample of the known Shakespeare to be tested in step six; 2) construct a random and sufficient group of plays from all of the other dramatists of the limited time frame and limited by all of the other constraints dictated by Shakespeare (e.g., tight limits on the genre and sub-genres);¹⁷ 3) perform all of the same steps that were done to obtain the Shakespeare sample; 4) determine if the selected style-markers actually differentiate between the known Shakespeare and the sample of other writers; 5) test the unknown sample; and 6) test the held-out sample of known Shakespeare. This necessary scenario is fraught with even more problems than obtaining a pure Shakespeare sample because, for instance, there were so many plays lost to time that we cannot get a random sample of all the plays written that fit the criteria.

Another type of control is the training set and cross validation methodology. There must be sufficient Shakespeare text after culling to employ these techniques – and I feel that there is not.

2.2.3 Eliminate or Control for the Variables Not being Tested

It is a basic principle in authorship attribution studies that the practitioner compare like to like. (Vickers 2011, 122)

One of the most important variables to be tested is genre. The genre of concern in this paper is drama; the sub-genre is drama written to be performed (vs. closet drama). Working down the sub-genre tree gives us the categories of comedy, tragedy, and history plays (there are others in this sub-genre but they will not be treated here). These three are how the First Folio is divided. What this means is that only the Shakespearean comedies should be used when testing if an unknown comedy should be placed in the Shakespeare canon. Now, taking only comedies and working further down the sub-genre tree, we come to prose vs. verse, rhymed verse vs. unrhymed, monologue vs. dialogue, plus there are the songs to be considered. Craig, Moscato and Rosso discuss the consequences of ‘single voice’ works and ‘dialogue’ (2009, 920). By only using rhymed verse as the known Shakespeare, for example, the demand for a tight genre approaches a *reductio ad absurdum*. But if practitioners go higher up the genre tree, they must show that their choice of style markers shows a consistent usage pattern across the lower sub-genres.

It has been shown in many studies that genre trumps authorship – there is a greater stylistic difference between one author in different genres than between two authors writing in the same genre. These studies were not on Shakespeare and may not generalize over to him, but the practitioners working on the Shakespeare canon must take the caveat seriously.

¹⁷ See Eder 2010 for a discussion of sample size.

There is no universal agreement on the chronology of the [Shakespeare] plays. (Hope and Whitmore 2014, 22)

The commonplace editorial concerns over a play's 'date of composition' which assumes a relatively limited amount of time during which a text was fully composed and after which it was merely transmitted and corrupted, is obviously problematic in this broader understanding of collaboration. (Masten 1997, 15)

In some cases we acknowledge that similarity in topic rather than authorship may be the best explanation for a close relationship between texts. (Arefin *et al.* 2014, 1)

Chronology is another important variable to be controlled. There can be no doubt that for the majority of authors, style changes over time. Hope and Whitmore speak of dividing the dramas into 'chronological periods', but even this larger division is not certain – they list 'one of several' possibilities (2014, 22). Rosso, Craig, and Moscato's study shows chronological separation between 1) early and middle comedies and two tragedies, and 2) history plays and mostly later comedies and tragedies (2009, 922). Jacqueline Mullender also talks about the linguistic differences in the later plays; she 'finds substantial evidence of increased syntactic complexity, and identifies significant linguistic differences between members of the wider groups of later plays' (2010, iii).¹⁸ Vickers talks about the problems for his tests caused by chronology: '[The tests are] quite reliable for ... third and fourth periods ... less successful, however, for the chronology of plays dated between 1595 and 1599' (2004, 106–107). A valid study of how Shakespeare's style changes over time is a *sine qua non* for setting time constraints on input texts – and such a study does not exist.

Outlier[s] may be explained by different genre, chronology may also be a controlling factor ... The poorest attribution [of Jonson's work is] a romantic comedy vs. his usual satire. (Budden *et al.* 2013, 10)

The point I would like to again emphasize is that these variables should be controlled for before the texts are determined.

2.2.4 Choice of Style Markers

Elizabethan drama is a genre in which authors are not immediately visible: they speak through their characters, who are individualized according to gender, age,

¹⁸ I am aware that I am on the horns of a dilemma. I argue that we do not know exactly what words Shakespeare wrote in the plays and yet I cite authors who assume that the writing is by Shakespeare. The point is that, even if Shakespeare wrote enough of the words to wash out any statistical significance of those words he did not write, there are enough other problems to keep the practitioner from doing attribution studies on Shakespeare's canon.

social class, and dramatic function. A simple computation of function words, however elaborately sifted by statistical procedures, may tell you something about the characters but cannot reliably indicate authorship. (Vickers 2009, 42)

Non-traditional authorship attribution study has turned its back on rare words, since their infrequency makes them useless for statistical purposes ... practitioners failed to realize that writers could and did borrow new and rare words from each other. (Vickers 2011, 123)

The punctuation in printed texts in most cases probably owes little to what Shakespeare wrote and is repeatedly misleading to the modern eye. (Jowett 1999, 72)

They [enclitic and proclitic microphrases developed by Marina Tarlinskaja 1987] are slow, manual, complex, require judgment, and are not easy to learn, replicate, or even describe completely. But, properly done, they are replicable enough and distinguishing enough to be one of the most powerful tools. (Elliott and Valenza 2004, 125)

The number of style markers in non-traditional authorship attribution moved from the thousands to the millions with the introduction of Docuscope, a text analysis program developed by David Kaufer's group at Carnegie Mellon University that introduced rhetoric strings as style markers (Butler *et al.* 2004). Hope and Whitmore use these strings as style markers and point out some of their shortcomings. The newest style marker is from Clarke (2013): Rare Collocation Profiling, or RCP. He identifies rare phrases and finds authors who share their use. Almost every practitioner's choice of style markers has met with negative criticism. It is only with a choice of as many different and statistically independent markers that can be shown to be consistent and constant across the selected sample of Shakespeare work and also be shown to exhibit different patterns in the authors of the control group can the selected markers be deemed valid.

2.2.5 *Other Choices*

Without going into great detail, there are many other choices facing the non-traditional practitioner working on Shakespeare, such as lemmatization, regularizing the spelling and grammar, disambiguating homographs, fixing 'obvious' printing mistakes, filling in lacunae. Elliott and Valenza say, in speaking about Jackson's work, 'He counted not just the word itself, but also some of its roots and kin' (2004, 120). Horton has a good discussion of the pros and cons of modernizing the text, such as spelling, hyphens, and contracted forms, and goes on to say, 'Nothing is to be gained from a hodge-podge of ancient and modern' (1987, 31). Horton stands alone in taking pains to explain everything he does in his study. He also stands alone in testing

alternative approaches, effectively doubling the work, such as the effect of lemmatizing on his non-lemmatized results.

Markup should also be considered. One of the reasons for using markup (e.g., the Text Encoding Initiative, or TEI) is that style markers such as the ratio of nouns to adjectives can be brought into the mixture. Craig and Hirsch discuss encoding e-texts and the need to update the encoding as the platforms evolve. They also discuss the TEI (2014, 17).

There are difficulties of course. An author can limit his style, vary it, imitate someone else to pose as that person, or write a parody so dependent on the original and so different from his own style ... [that it] is more difficult to discern. (Craig and Kinney 2009, 9)

2.2.6 *Statistical Tests*

The Morton-Merriam method identified *Edmund Ironside* as the work of Robert Greene, not Shakespeare, at odds announced as 890 million million to 1. This calculation impressed many readers, including at least one stylometrist (D.F. Foster). But even Dr. Smith has denounced such investigators as 'mesmerized by their arithmetic at the expense of their critical faculties' ... Not only do Morton and Smith contradict each another, they are both contradicted by other stylometrists [sic] such as Brainerd and Slater, each of whom inferred from his own separate statistical system that *Ironside* was in fact Shakespeare-compatible. (Love 2002, 155, quoting Sams 1994, 471-472)

All statistical work, however sophisticated, however crude, is inevitably hedged around with caveats and disclaimers, and this is both reasonable and necessary. But we must make sure ... that the sum of the qualifiers is not greater than the usefulness of the statistics. (Eliot 2002, 286)

The most important concept for non-statisticians to grasp is the fact that statistical tests most often come with assumptions. The most common assumptions are randomness and independence, yet Shakespeare's words are neither random nor independent. Budden, Craig, Marsden and Moscato consider 'language and the potential of words as an abstract chaotic system', but they go on to say that 'authors are required to adhere to the grammatical and structural rules dictated by a written language' (Budden *et al.* 2013, 1). Vickers discusses a few problems with Principal Component Analysis (PCA) in non-traditional authorship attribution as pointed out by Maciej Eder in a private communication with Vickers and also by Patrick Juola in his 2006 article, 'Authorship Attribution' (Vickers 2011). Burrows defends the technique.

Another concept to worry about is cherry picking – trying various input texts or sample sizes or trying various tests and selecting the ones that work.¹⁹

¹⁹ For a discussion of this, see Rudman 2003.

Oakes critiques several statistical techniques used by practitioners working on the Shakespeare canon.

Three of the Craig collaborations exhibit some of the most sophisticated statistics used in Shakespeare attribution studies and just the titles give the statistically naïve readers pause:

- 1) 'An Information Theoretic Clustering Approach for Unveiling Authorship Affinities in Shakespeare Era Plays and Poems' (Arefin *et al.* 2014)
- 2) 'Language Individuation and Marker Words: Shakespeare and His Maxwell's Demon' (Budden *et al.* 2013)²⁰
- 3) 'Shakespeare and Other English Renaissance Authors as Characterized by Information Theory Complexity Quantifiers' (Craig, Moscato and Rosso 2009)

But keep in mind that statistics do not override the need for all of the other text-controlled variables. In talking about Morton's methodology, Taylor states: 'Whatever the theoretical usefulness of such traits in distinguishing authors, statistical analysis proves nothing when based upon masses of unreliable data ("garbage in, garbage out"). (1987b, 80)

There are very few statisticians working on the Shakespeare canon; Valenza is one. This causes many potential problems with practitioners adopting (and usually not adapting) statistical tests that were developed for different applications.

2.2.7 *Sample Selection and Size*

Various studies have clearly shown that the result of an authorship attribution method can be affected by parameters such as training corpus size, test corpus size, lengths of the texts (certain methods work effectively in the case of long texts but not well on short or very short texts), number of candidate authors, and distribution of the training corpus over the authors. (El Manar El Bouanani and Kassou 2014, 26)

There are different times in an attribution study when the practitioner is faced with sample selection and size. I spoke above about the random selection of non-Shakespearean writers for a control group. Unfortunately no practitioner does this. When they do use a control group they use a sample of convenience – whatever is at hand and easily used electronically. This, of course, invalidates 'randomness'.

Another sample selection is the size of the text blocks used in the analysis programs; for example, is the number of 'and' per 500-word block more meaningful than the number of 'and' in a 1,000-word block? Practitioner,

²⁰ For what it is worth: Maxwell's Demon is an imaginary creature created by the mathematician James Maxwell to contradict the second law of thermodynamics.

beware of experimental bias!²¹ There are other decisions to be made on sample size, such as 1) how large a sample of known Shakespeare do we need to do a valid study, and 2) how many words must the unknown work have in order to do a valid study? The answers to these are not easily determined.

2.2.8 *Treatment of Errors*

There are different kinds of errors: mistakes, data errors, statistical errors, and systematic errors. Every practitioner should be aware of all of the types of errors and how to handle them. A good starting point to understand errors is Yardley Beers' book (1958). Eder's paper on systematic errors in non-traditional authorship attribution is also a must-read (2012). Lancashire discusses the fact that probabilistic methods predict errors (2002). The data errors in e-texts are constantly being corrected. This makes the date that a practitioner accesses the text important. And we can only hope that the repository keeps such correction information on file. I have not seen one paper on the authorship of the Shakespeare canon that has an adequate treatment of errors.

2.2.9 *Analyze and Interpret Results*

One of the final steps of any study is to analyze and interpret the results. Again, do not be guided by experimental bias, as in 'my analysis does not give the right answer maybe because of some variable I failed to control'. It is dangerous for a non-statistician to try to explain and interpret statistical results, to try to explain the probabilities as probabilities and not percentages.

3. *Conclusion*

Whatever the future may bring, at the present time, the discipline [non-traditional authorship attribution] remains in flux. (Vickers 2011, 115)

I have studied, admired, and enjoyed the critical editions of Shakespeare's dramatic works through high school, college, graduate school, and in my professional and social life. The medium of drama does not demand a single author who penned every word. I believe we will never be able to know exactly which of the words in the dramas were written by William Shakespeare.

Non-traditional authorship attribution practitioners working on the canon of William Shakespeare are faced with serious – even fatal – problems

²¹ Do not go into a study to prove that Shakespeare wrote a newly uncovered work. Rather test to see who wrote the work and use Shakespeare as one of the candidates. The temptation is great because fame awaits.

to try to overcome. The most serious is that, in my opinion, we do not have sufficient texts to do a study. I feel that I have shown above that after the practitioner corrects for genre, chronology, collaboration, interpolation, and revision (to name the most important points) there are insufficient texts of what we are sure were written by Shakespeare and only Shakespeare.

Alternatively, one may accept the canon as a whole, hoping that its content of works by other authors is small enough not to affect broad results. (Love 2002, 197)

While pointing out some of the many problems and conflicts between the various non-traditional practitioners of Shakespearean authorship, I had the uneasy feeling that I might have finally united the field against a common enemy!

A little Learning is a dang'rous Thing;
 Drink deep, or taste not the Pierian Spring:
 There shallow Draughts intoxicate the Brain,
 And drinking largely sobers us again. (Pope 1711, 14)

Is this meant as a warning to the Shakespeare scholars who venture into the realm of non-traditional authorship attribution studies – or something I should heed venturing onto their turf? I think there is enough caution to go around. I also feel there cannot be a valid non-traditional authorship attribution study of the Shakespeare canon using the present day state of the science.²²

Works Cited

- Arefin A.S., Hugh Craig, Pablo Moscato, Carlos Riveros, and Renato Vimieiro (2014), 'An Information Theoretic Clustering Approach for Unveiling Authorship Affinities in Shakespearean Era Plays and Poems', *PLOS ONE* 9.10, 1-12 (E111445), <<http://www.plosone.org>>, accessed 30 October 2014.
- Bate Jonathan (2013), 'General Introduction', in J. Bate, E. Rasmussen *et al.*, eds, *William Shakespeare and Others: Collaborative Plays*, The RSC Shakespeare, Basingstoke, Palgrave Macmillan, 9-30.
- Beers Yardley (1958), *Introduction to the Theory of Errors*, Cambridge, Addison Wesley.
- Best Michael (2007), 'Shakespeare and the Electronic Text', in A. Murphy, ed., *A Concise Companion to Shakespeare and Text*, Oxford, Blackwell Publishing, 145-161.

²² When I was asked to write this paper I hesitated – it seemed like a no-win situation. I realized that the points to be made could not be supported with enough facts and examples from extant studies to make the points as well as I wanted and still fit within the editorial constraints. Hopefully, the monograph will make the case clearer and stronger.

- Blaney P.W.M. (1991), *The First Folio of Shakespeare*, Washington D.C., Folger Shakespeare Library, electronic edition in *William Shakespeare London 1623, Comedies, Histories, & Tragedies*, bibliographic items, 1[- GB]18, <<http://hamnet.folger.edu/other/folio/Shaf1B.pdf>>, accessed 21 June 2014.
- Budden David, Hugh Craig, John Marsden, and Pablo Moscato (2013), 'Language Individuation and Marker Words: Shakespeare and His Maxwell's Demon', *PLoS ONE* 8(6): e66813, <<http://dx.doi.org/10.1371/journal.pone.0066813>>, accessed 30 June 2013.
- Bullough Geoffrey (1957), *Narrative and Dramatic Sources of Shakespeare*, vol. I, London, Routledge and Kegan Paul, 8 vols.
- Bullough Geoffrey (1962), *Narrative and Dramatic Sources of Shakespeare*, vol. IV, London, Routledge and Kegan Paul, 8 vols.
- Bullough Geoffrey (1975), *Narrative and Dramatic Sources of Shakespeare*, vol. VIII, New York, Columbia University Press, 8 vols.
- Burrows John (2008 [2007]), 'A Reply to Joseph Rudman's "Riposte" ', *Script & Print* 31, 4, 220-229.
- Burrows John (2012), 'A Second Opinion on "Shakespeare and Authorship Studies in the Twenty-First Century" ', *Shakespeare Quarterly* 63, 3, 355-392.
- Butler Brian, Jeff Collins, David Kaufer, and Ishizaki Suguru (2004), *The Power of Words: Unveiling the Speaker and Writer's Hidden Craft*, London, Routledge.
- Clarke Barry (2013), 'A Linguistic Analysis of Francis Bacon's Contributions to Three Shakespeare Plays: *The Comedy of Errors*, *Love's Labour's Lost*, and *The Tempest*.' Ph.D. Dissertation, London, Brunel University.
- Craig Hugh and A.F. Kinney (2009), 'Introduction', in H. Craig, A.F. Kinney, eds, *Shakespeare, Computers, and the Mystery of Authorship*, Cambridge, Cambridge University Press, 1-14.
- Craig Hugh and B.D Hirsch (2014), ' "Mingled Yarn": The State of Computing in Shakespeare 2.0', in T. Bishop, A. Huang, gen. eds, H. Craig, B.D. Hirsch, guest eds, *The Shakespearean International Yearbook* 14, Aldershot, Ashgate, 3-35, preprint edition, courtesy of publisher.
- Craig Hugh, Pablo Moscato and O.A. Rosso (2009), 'Shakespeare and Other English Renaissance Authors as Characterized by Information Theory Complexity Quantifiers', *Physica A*, 388, 6, 916-926.
- Eder Maciej (2010), 'Does Size Matter? Authorship Attribution, Small Samples, Big Problems', given at ADHO 2010, paper courtesy of author.
- Eder Maciej (2012), 'Mind Your Corpus: Systematic Errors in Authorship Attribution', paper delivered at *Digital Humanities 2012*, abstract in *Digital Humanities 2012: Conference Abstracts*, Hamburg, Hamburg University Press, 181-185, full paper courtesy of author.
- Efron Bradley and Ronald Thisted (1976), 'Estimating the Number of Unseen Species: How Many Words did Shakespeare Know?', *Biometrika* 63, 3, 435-447.
- Eliot Simon (2002), 'Very Necessary but Not Quite Sufficient: A Personal View of Quantitative Analysis in Book History', in E. Greenspan, J. Rose, eds, *Book History*, vol. V, University Park, Pennsylvania State University Press, 283-293.
- Elliott W.E.Y and Robert Valenza (1996), 'And Then There Were None: Winnowing the Shakespeare Claimants', *Computers and the Humanities* 30, 3, 191-245, with revised appendices 1999; courtesy of the authors.

- Elliott W.E.Y and Robert Valenza (2004), 'Did Shakespeare Write *A Lover's Complaint*? The Jackson Ascription Revisited', in B. Boyd, ed., *Words That Count: Essays on Early Modern Authorship in Honor of MacDonald P. Jackson*, Newark, University of Delaware Press, 117-140.
- El Manar El Bouanani Sara and Ismail Kassou (2014), 'Authorship Analysis Studies: A Survey', *International Journal of Computer Applications* 86, 12, 22-29.
- Farmer A.B. (2002), 'Shakespeare and the New Textualism', in W.R. Elton and J.M. Mucciolo, eds, *The Shakespearean International Yearbook*, vol. II, *Where are We Now in Shakespearean Studies?*, Aldershot, Ashgate Publishing, 158-179.
- Groom Nick (2003 [2002]), 'Forgery, Plagiarism, Imitation, Peglegger', in P. Kewes, ed., *Plagiarism in Early Modern England*, New York, Palgrave Macmillan, 74-89.
- Hinman Charlton (1963), *The Printing and Proof-Reading of the First Folio of Shakespeare*, Oxford, Clarendon Press, 2 vols.
- Hope Jonathan (1994), *The Authorship of Shakespeare's Plays: A Socio-Linguistic Study*, Cambridge, Cambridge University Press.
- Hope Jonathan and Michael Whitmore (2014), 'Quantification and the Language of Later Shakespeare', *Actes des Congrès de la Société Française Shakespeare* 31, 123-149.
- Horton T.B. (1987), 'The Effectiveness of the Stylometry of Function Words in Discriminating Between Shakespeare and Fletcher', Ph.D. thesis (CST-48-87), University of Edinburgh.
- Ioppolo Grace (2006), *Dramatists and their Manuscripts in the Age of Jonson, Middleton and Heywood: Authorship, Authority and the Playhouse*, London, Routledge.
- Jackson MacDonald P. (2003), *Defining Shakespeare: Pericles as Test Case*, Oxford, Oxford University Press.
- Jackson MacDonald P. (2014), *Determining the Shakespeare Canon: 'Arden of Faversham' and 'A Lover's Complaint'*, Oxford, Oxford University Press.
- Jowett John (1999), 'After Oxford: Recent Developments in Textual Studies', in W.R. Elton and J.M. Mucciolo, eds, *The Shakespearean International Yearbook Vol. 1: Where are We Now in Shakespearean Studies?*, Aldershot, Ashgate Publishing, 65-86.
- Juola Patrick (2006), 'Authorship Attribution', *Foundations and Trends in Information Retrieval* 1, 3, 233-334; this paper carries a copyright of 2008 and there are references beyond the 2006 publication date.
- Kositsky Lynn and R.A. Stritmatter (2013), *On the Date, Sources and Design of Shakespeare's 'The Tempest'*, Jefferson, McFarland & Company, Inc.
- Lancashire Ian (2002), 'The State of Computing in Shakespeare', in W.R. Elton and J.M. Mucciolo, eds, *The Shakespearean International Yearbook*, vol. II: *Where are We Now in Shakespearean Studies?*, Aldershot, Ashgate Publishing, 89-110.
- Love Harold (2002), *Attributing Authorship: An Introduction*, Cambridge, Cambridge University Press.
- Maguire Laurie and Emma Smith (2012), 'Many Hands. A New Shakespeare Collaboration?', *Times Literary Supplement*, 20 April, 13-15.
- Masten Jeffrey (1997), *Textual Intercourse: Collaboration, Authorship, and Sexualities in Renaissance Drama*, Cambridge, Cambridge University Press.
- McCarty Willard (2011), 'Editorial', *Interdisciplinary Science Reviews* 36, 4, 273-275.
- Muir Kenneth (1978), *The Sources of Shakespeare's Plays*, New Haven, Yale University Press.

- Mullender Jacqueline (2010), 'Shakespeare's Late Syntax: A Comparative Analysis of *Which* Relativisation in the Dramatic Works', Ph.D. Thesis, University of Birmingham.
- Oakes M.P. (2014), *Literary Detective Work on the Computer*, Amsterdam, John Benjamins Publishing Company.
- Pope Alexander (1711), *An Essay on Criticism*, London.
- Rasmussen Eric (1997), Review of Jonathan Hope, *The Authorship of Shakespeare's Plays*, *Modern Philology* 95, 1, 109-113.
- Reisz Matthew (2014), 'Shakespeare Scholar Disputes Decision Reversal by Journal', *Times Higher Education*, 11 September.
- Rudman Joseph (1998), 'The State of Authorship Attribution Studies: Some Problems and Solutions', *Computers and the Humanities* 31, 4, 351-365.
- Rudman Joseph (2003), 'Cherry Picking in Non-Traditional Authorship Attribution Studies', *Chance* 16, 2, 26-32.
- Rudman Joseph (2005), 'Unediting, De-editing, and Editing in Non-traditional Authorship Attribution Studies: With an Emphasis on the Canon of Daniel Defoe', *Papers of the Bibliographical Society of America* 99, 1, 5-36.
- Rudman Joseph (2006), 'Authorship Attribution: Statistical and Computational Methods', in K. Brown, ed., *Encyclopedia of Language and Linguistics*, second edition, London, Elsevier, 611-617.
- Rudman Joseph (2011), 'Stylometrics', in P.C. Hogan, ed., *Cambridge Encyclopedia of the Language Sciences*, Cambridge, Cambridge University Press, 817-819.
- Rudman Joseph (2012), 'The State of Non-Traditional Authorship Attribution—2012: Some Problems and Solutions', *English Studies* 93, 3, 259-274.
- Sams Eric (1994), 'Edmund Ironside and Stylometry', *Notes and Queries* 239, 469-472.
- [SAS] (2005), 'Concepts of Experimental Design', *A SAS White Paper*, 2005, <<http://support.sas.com/resources/papers/sixsigma1.pdf>>, accessed 10 November 2014.
- Satin Joseph (1966), *Shakespeare and His Sources*, Boston, Houghton Mifflin.
- Schoenbaum Samuel (1966), 'Internal Evidence and the Attribution of Elizabethan Plays', in D.V. Erdman, E.G. Fogel, eds, *Evidence for Authorship: Essays on Problems of Attribution*, Ithaca, Cornell University Press, 188-203.
- Scott Gray (2006), 'Signifying Nothing? A Secondary Analysis of the Claremont Authorship Debates', *Early Modern Literary Studies* 12, 2, 1-50, <<http://purl.oclc.org/emls/12-2/scotsog2.htm>>, accessed 13 November 2014.
- Stern Tiffany (2011), '“The Forgery of Some Modern Author”? Theobald's Shakespeare and Cardenio's *Double Falsehood*', *Shakespeare Quarterly* 62, 4, 555-593.
- Tarlinskaja Marina (1987), *Shakespeare's Verse: Iambic Pentameter and the Poet's Idiosyncrasies*, New York, Peter Lang.
- Taylor Gary (1987a), 'General Introduction', in S. Wells, G. Taylor, J. Jowett and W. Montgomery, eds, *William Shakespeare: A Textual Companion*, Oxford, Clarendon Press, 1-68.
- Taylor Gary (1987b), 'The Canon and Chronology of Shakespeare's Plays', in S. Wells, G. Taylor, J. Jowett and W. Montgomery, eds, *William Shakespeare: A Textual Companion*, Oxford, Clarendon Press, 69-144.
- Taylor Gary (2016), 'Fake Shakespeare', *Journal of Early Modern Studies* 5, 353-379.
- Tomimura Noritaka (2009), 'The Use of Song in Shakespeare's Plays', in M. Hori, S. Kumamoto and T. Tabata, eds, *Stylistic Studies of Literature*, Bern, Peter Lang, 161-187.

- Vickers Brian (2002), *Shakespeare, Co-Author: A Historical Study of Five Collaborative Plays*, Oxford, Oxford University Press.
- Vickers Brian (2004), 'The *Troublesome Raigne*, George Peele, and the Date of *King John*', in B. Boyde, ed., *Words That Count: Essays on Early Modern Authorship in Honor of MacDonald P. Jackson*, Newark, University of Delaware Press, 78-116.
- Vickers Brian (2007), 'Incomplete Shakespeare: Or, Denying Coauthorship in *I Henry VI*', *Shakespeare Quarterly* 58, 3, 311-352.
- Vickers Brian (2009), 'The Marriage of Philology and Informatics', *British Academy Review* 14, 41-44.
- Vickers Brian (2011), 'Shakespeare and Authorship Studies in the Twenty-First Century', review essay, *Shakespeare Quarterly* 62, 1, 106-142.
- Vickers Brian and Marcus Dahl (2012), '*All's Well That Ends Well*: An Attribution Refuted', *Times Literary Supplement*, 12 May, 14-15.