THE PROOF IS IN THE PEMBROKE

A Stylometric Comparison of the Works of Shakespeare with 12 Works by 8 Elizabethan Authors

George Warren

S TATISTICAL studies of words used infrequently by Shakespeare have become increasingly important in the scholarly community. Eliot Slater's Rare Word Test of 1977 is often quoted in scholarly books and essays like *The Oxford Textual Companion* as proof of the chronological relationships between various works by Shakespeare (Taylor 109-34). Recently, rare words analyses by MacDonald P. Jackson and the Hiatts have been used to establish a chronology for the writing of the *Sonnets* (103-05). Rare words have been used to suggest or confirm dates of composition of certain works by Shakespeare by referencing texts of other authors which used the same rare words.[1] I adopted this idea and constructed a set of rare words in order, first, to test Slater's results, and second, to compare rarely used words of Shakespeare to twelve works by other authors: Christopher Marlowe's *Tamburlaine I, Tamburlaine II* and *Edward II,* Robert Greene's *Friar Bacon and Friar Bungay* and *James IV,* Mary [Sidney Countess of] Pembroke's *Antonius* and Psalms 95-150, Thomas Nashe's *Summer's Last Will and Testament,* Anthony Munday's *John a Kent,* Thomas Kyd's *The Spanish Tragedy*, Thomas Middleton's *A Trick to Catch the Old One*, and John Fletcher's *The Faithful Shepherdess.*

Procedure

Outside of spreadsheet calculations, most of my study was done "by hand." If Oxfordians deem it worthwhile, hopefully this study can be replicated and expanded by computer studies. I drew my list of rare words from Bartlett's *Concordance*, which removed *The Two Noble Kinsmen* and *Edward III* from analysis. I had intended to include all words used between two and ten times in Shakespeare's plays, but this proved to be too cumbersome, so I eventually settled on a study of roughly five thousand words used between two and ten times selected at random from Bartlett's compilation.

Assembling the data

Basing my study on this set of rarely used words, I constructed a list of Shakespeare's plays along the X-axis versus the same thirty-seven plays along the Y-axis. I then tabulated the number of times each play shared a rare word with the other plays. For example, the word "mitigate" was used by Shakespeare three times according to Bartlett; once each in *Richard III, 1 Henry VI,* and *Merchant of Venice.* This created usage hits for *Richard III* against *1 Henry VI* and MOV, for *1 Henry VI* against *Richard III* and *MOV,* and for *MOV* against *Richard III* and *1 Henry VI.* These I recorded on the matrix and then followed the same procedure for the remaining five thousand words, adding all the "hits" for the various 5000 rare words to the same matrix.[2] This procedure

established the total number of rare words shared between each pairing of plays.

Due to the varying lengths of the plays, the resulting data was skewed so that the raw data suggested that thirty-four of the thirty-six plays other than *Hamlet* were written near the time of *Hamlet's* composition. This required a normalization of the data by line count.[3]

I compared the resulting normalized data with Slater's "Rare Word Analysis" as cited in *The Oxford Textual Companion*. Although I didn't duplicate his results, I did confirm much of what he'd shown. For example, the plays Slater found most similar to *1 Henry VI* in the use of rarely used words were, in order, *3 Henry VI, Richard III, Titus Andronicus* and *The Two Gentlemen of Verona*. My test for *1 Henry VI* suggested the plays closest to it in the use of rare words were, in order, *2 Henry VI, 3 Henry VI, Titus*, and *Richard III*. A second example: for *Richard II*, Slater's test showed the following to be closest, *Richard III, Titus Andronicus*, the *Henry VIs* and *King John*. My test showed *2 Henry VI, 1 Henry VI, Richard III* and *King John*. And a third example: for *Troilus and Cressida*, Slater's test showed those most nearly related to be *Hamlet, Henry V* and *Macbeth*. Mine showed *Hamlet, Henry V, All's Well That Ends Well* and *Macbeth*.

My test also showed certain close relationships not shown in Slater's test. For instance, in my test the play with rare word usage most similar to *The Taming of the Shrew* was *1 Henry IV*.

Comparison of Shakespeare with other plays

In order to compare the Shakespeare canon with other writers I calculated the average number of hits for each play. This normalized data was headed by *Hamlet* with an average of seventy-seven hits per play (high score, *Lear*: 112; low score, *Two Gents*: 48). The lowest average number of normalized hits was *Two Gentlemen of Verona* with 42 (high score: *Taming of the Shrew*: 61, low score: *Antony and Cleopatra*: 25). The average number of shared rare words (77 in *Hamlet*, 42 in *Gents*) was compared with the twelve works by other authors. If the others' scores against the canon exceeded this average by 20 or more shared rare words, I thought they deserved further study to determine possible influence of one work on the other. Scores of 97 hits, against *Hamlet*, or 62 against *Gents*, matched that +20 standard and these unusually high scores were interesting enough to provoke a closer look.

I then took the complete word lists for the twelve works by other writers mentioned above

and compared them with the list of rare words used by Shakespeare, locating each word that appeared on both lists (ie. words used both by Shakespeare and a second writer) in Bartlett's *Concordance* and noting in which plays Shakespeare had used this word. I credited the work in which the word appeared with one hit against each Shakespearian play which used it. For example, if Marlowe used the word "mitigate" in *Edward II, Edward II* would be credited with one word likeness each against *Richard III, 1 Henry VI,* and *Merchant of Venice.*

I then totaled the number of hits against each play and normalized by line count. I compared the resulting number of normalized hits against each of Shakespeare's plays with the average number of normalized hits the Shakespeare matrix had generated. For example, on the Shakespeare matrix the average number of normalized hits for *Titus Andronicus* is 58. Marlowe's *Edward II* shared 94 rare words with *Titus,* which was normalized by line count to 100. Subtracting the average for *Titus* of 58 from *Edward II's* normalized total of 100 yielded a score of 42 for *Edward II* against *Titus.* I then compared each work by the other writers in this same fashion with each of Shakespeare's plays and made a list of these comparisons for each writer, ordered from highest magnitude to lowest. For example, the top five scores for Marlowe's *Edward II* were *Titus Andronicus*: 42, *3 Henry VI*: 31, *1 Henry VI*: 29, *Richard II*: 21, and *2 Henry VI*: 18. (For each of these comparisons an average score would have been 0.)

A summary of these comparisons against the twelve works by other authors revealed that of the 444 comparisons (12 works times 37 works by Shakespeare), 391 or 88.1% scored below twenty; 30 or 6.7% scored 20 to 29.9; 15 or 3.4% scored 30 to 39.9; and 8 or 1.8% scored 40 or higher. Interestingly, Marlowe's *Edward II* compared with *Titus Andronicus* achieved one of the top eight scores. This suggested that scores of 30 or above should be looked at more closely as they might be exhibiting influence between the two plays, authorship of two or more writers, or later changes and additions by an editor.

Scene by scene analysis

For those works that achieved the high scores and for a great number of those that did not, I then conducted a second study: a scene by scene analysis of each of Shakespeare's plays to see if the shared rare words clustered in particular scenes. Based on raw data before normalization, two plays by Shakespeare on average would share a rarely used word every 45 lines or about every 43 lines with repetitions of the rare words. For instance, when I compared the Shakespeare canon with the Countess of Pembroke's translation of Psalms 95-150, these revealed a Shakespearean rarely-used word once per 30 lines or more in 65.4% of the total scenes in Shakespeare, which means that she used an infrequently-used word per every 19.9 or less lines in only 12.7% of the total scenes in Shakespeare or in only 6.8% of the scenes in Shakespeare of a hundred lines or more.

An average score for Shakespeare himself would be one rare word in 43 lines and here is Pembroke using one of his rare words in 20 lines or less, rare words found in only 12.7% of all the

scenes in Shakespeare. Which is more likely: that Shakespeare would suddenly, for an entire scene, start using a lot of his rarely used words together? or that someone known to use these words—and maybe not just in rare instances—would have used them? For that 6.8% of the long scenes it seems worthwhile to take a closer look at Pembroke as possible author or editor.

For the twelve works studied by other writers I decided that scores of less than 20 lines per rare word in a scene, if these scores appeared four or more times in a play—that is, they appeared in at least four scenes—should be looked at as a possible indicator of influence.

Relevance

The possibility of this study's numbers being purely random and without significance needs to be considered. Marlowe's *Tamburlaine* I (1587) generated the following numbers for the four plays with the highest scores above average: *1 Henry VI*: 60, *3 Henry VI*: 30, *Titus Andronicus*: 29, and *2 Henry VI*: 27. There are at least three reasons why these numbers cannot be random spikes. Spikes of 40 or more were quite rare in this study, and spikes of sixty or more occurred only twice. (*Tamburlaine I* scored 60 against *1H6* and *Last Will* scored 68 against *Titus*.) A spike of such magnitude would be extremely hard to generate randomly in this kind of rare word study. Secondly, on the list of plays ranked by average number of lines per rare word, two of these four plays had relatively high ratios: *1 Henry VI* with one rare word expected every 38.3 lines is second on this list; *2 Henry VI* with one per 40.9 lines is fourth; but *Titus Andronicus* with one per 45.1 lines is eighteenth and *3 Henry VI* with one per 48.3 lines is twenty-seventh of the 37 plays.

You would expect a play with one rare word every 38.3 lines (like *1H6*) to have a higher number of normalized hits than one with one rare word per 45 lines, if nothing else intervenes. In a random selection, this would be expected to occur.[5]

Remember that in Shakespeare himself, those numbers of lines per rare word are an average of all comparisons of two plays. In *Gents* he may use the words that appear in 1 & 4, while in *Hamlet* he uses only those in 2, and in *1H4* he uses only those that appear in 1 & 3. Thus, on average, he was using those in 1 the most. When you combine all 37 plays, #1 may be at the top overall, but not in *Hamlet*, and not in *Marlowe*.

Finally, there are the scores achieved by the other two plays by Marlowe considered in this study. *Tamburlaine II* (1588) had the following four highest scores: *1 Henry VI*: 47, *3 Henry VI*: 40, *Titus Andronicus*: 33, and *2 Henry VI*: 29. For *Tamburlaine II*, not only are the four plays identical to the four scoring highest for *Tamburlaine I*, they are even in the same order. In 1592, four years after *Tamburlaine II*, Marlowe composed *Edward II*. The highest scores for *Edward II* in this study were: *Titus Andronicus*: 42, *3 Henry VI*: 31, *1 Henry VI*: 29, *Richard II*: 21, and *2 Henry VI*: 18. The odds against the four plays appearing randomly on all three lists of highest scores in the top five is overwhelming. From this it can be seen that this rare words study is showing some sort of connection between four of Shakespeare's plays and the three plays by Marlowe.

Nor can the high numbers generated in the comparison of Mary Pembroke's translation of Garnier's *Marc Antoine* with the Shakespeare canon be merely random spikes. For Pembroke's

Antonius, translated by November 1590, the plays that finished in the 6th, 7th, 9th, and 10th places on her list of highest scores were *Julius Caesar, Antony and Cleopatra, Coriolanus,* and *Timon of Athens*. These plays are thought by orthodox scholars to have been composed by Shakespeare in 1599, 1606, 1608, and 1605 (Taylor 109-33) respectively and yet Mary Sidney's 1590 translation scored high against all of them!

What these four plays have in common is that they are the four plays written by Shakespeare with Plutarch as their primary source. Garnier's *Marc Antoine* was also based on Plutarch and both Pembroke and Shakespeare used North's translation of Plutarch in the creation of their plays (Hannay 140). To reiterate: the four plays where Shakespeare was most dependent on Plutarch all appear in the ten highest scores achieved by Pembroke's translation of a play based on Plutarch. Rather than four random spikes, these numbers prove that in this case, the rare words test is measuring Pembroke's and Shakespeare's dependence on a particular source: Plutarch. In addition to this, the plays that placed just higher than these four Plutarch-based Shakespeare dramas, those in the 3rd, 4th, and 5th positions, were all known to have been performed by Pembroke's Men, whose patron was the Earl of Pembroke, the Countess's husband.[4] *2 Henry VI* finished 3rd, *Titus Andronicus* finished fourth, and *3 Henry VI* was 5th.

This rare words test, therefore, is not simply generating a series of random spikes. It is showing connections between these authors and specific works in the Shakespeare canon. The question is, what do these connections signify?

Meaning

For Pembroke's high scores on *Antonius* when compared with Shakespeare's Plutarch-based plays, the meaning is obvious: the test has measured a dependence of two playwrights on the same source. However, in the case of the three plays by Marlowe, although the test shows a similar connection, it cannot be showing the same type of influence since there are a variety of primary sources for the *Tamburlaines*, *Edward II*, *Titus Andronicus*, and the *Henry VIs*. Perhaps the rare words study shows that *Tamburlaine I*, written in 1587, was dependent on *Titus* and the *Henry VIs*. Orthodox scholars, of course, see this the other way round, as Shakespeare indebted to *Tamburlaine* for these four plays which they hold were written in 1591 and 1592 (Cox 2, 104-5; Forker 159). If Shakespeare was just starting out, what was more natural than to imitate the successful Marlowe?

However, orthodox scholars shift gears when it comes to *Edward II*. Because of the success Shakespeare had supposedly achieved with his history plays, the *Henry VIs*, they believe it was Marlowe who borrowed from Shakespeare when he composed his own first history play in 1592. With *Richard II* they reverse the influence once again, since *Richard*, thought to have been composed in 1595, could not possibly have been a source for *Edward II*, which was written in 1592 (159-164). It could be claimed that my study confirms this since *Richard II* was ranked fourth against *Edward II* and hadn't registered in the top four against the *Tamburlaine*s. But it could also go the other way, since the same data, showing some obvious connection, could be interpreted as showing that Marlowe was influenced by Shakespeare for all three of Marlowe's plays. There is another pos-

sibility: since all four of these plays by Shakespeare are thought by orthodox scholars to be collaborative works, Marlowe may have had a hand in writing them.[5]

A score of thirty or more in the normalized section of the study might indicate participation in the writing of a play credited to Shakespeare. But Fletcher's *The Faithful Shepherdess* scored 35 against Shakespeare's *A Midsummer Night's Dream* although the dates make it impossible for Fletcher to have contributed to that play.

Antonius

Since the plays ranked 6th, 7th, 9th, and 10th reflect Pembroke's use of a specific source, what do her higher scores indicate? Those ranked immediately above Shakespeare's Plutarch-based plays in 3rd, 4th, and 5th places are, as we have seen, *2 Henry VI*, *Titus Andronicus*, and *3 Henry VI*—all known to have been played by Pembroke's Men.[6] Orthodox scholars date the composition of these plays to 1591, 1592, and 1591 respectively. Until recently it was thought they couldn't have been performed before the formation of Pembroke's Men in the second half of 1592. More recently Andrew Gurr has challenged this by moving the formation of this company back to mid-1591 (267).

Mary's *Antonius* is known to have been completed in November 1590 and published in 1592 (Bullough 229, 358). The dates of composition for *2 Henry VI*, *3 Henry VI* and *Titus Andronicus* given by orthodox scholars are placed between the completion of Pembroke's *Antonius* and its publication. If the orthodox scholars are wrong and Shakespeare's plays were known to Pembroke earlier than the orthodox dates, the influence would go from Shakespeare to Pembroke. If the orthodox dates are correct, despite finishing higher on her list of rare words, Pembroke could not have used these works as sources for *Antonius* and the connection must be influence in the other direction, with Shakespeare reflecting its language in his own plays as he had done, we are told, with Marlowe's *Tamburlaine*s. This would mean the new kid on the block in 1591 had access to a manuscript written by the Countess of Pembroke (her play wouldn't be published for another year); ie. that the unproven youth from Stratford had access to the private papers of the most important literary patroness in England. And this, if we accept the orthodox chronology, was just the beginning of his borrowings from manuscripts known to be in her possession.[7] If the Shakespeare described by orthodox scholars did in fact borrow from Pembroke's manuscript of *Antonius*, then the Countess must have played a major if unrecognized role in the life of the great Shakespeare.

Still, orthodox scholars might look at the words Pembroke used that match with Shakespeare and still conclude that Shakespeare was not influenced by her vocabulary and had not used her manuscript. Then what would have caused the spikes for these three plays in the data developed from *Antonius*? There are two possibilities. First, that all three were coincidentally just random spikes; it was nothing more than coincidence that three plays performed by Pembroke's Men happened to score highly. Or, the three plays preceded Mary's translation and she was sufficiently familiar with them to have thoroughly absorbed their use of language.

To me the orthodox view of Shakespeare seems a stretch. If there is a different and simpler way of looking at this same data, why go through all these convolutions to achieve the less believable?

At minimum, we should adjust the dates of composition for these plays by Shakespeare to earlier dates proposed by scholars like E.A.J. Honigmann, who lists the dates of composition for *Titus Andronicus* as 1586, *1H6* as 1588, *2H6* as 1589, and *3H6* as 1590 (128). Thus *2 Henry VI, 3 Henry VI,* and *Titus Andronicus* were written prior to the completion of Pembroke's *Antonius* in November 1590. This would still allow the orthodox scholars to maintain their ideas about the influences of Shakespeare and Marlowe on each other. But even this relationship would be simpler and easier to believe if those three plays were written before the *Tamburlaine*s, not after.

There were two plays that scored even higher than those noted above in the comparison of Antonius with Shakespeare. *Pericles* was second at 21 and *1 Henry VI* was at the top with 37. Orthodox scholars hold (not very convincingly) that *Pericles* was a 1607 collaboration between Shakespeare and Wilkins (557). It is obvious that, in the orthodox scenario, Pembroke could not have used *Pericles* (1607) as a source for *Antonius* (1590). And why did *1H6* score so much higher than the other plays?

To get a second look at the vocabulary Pembroke employed while writing verse, I turned to her translation of the Psalms. I divided them into two parts, each of approximately equal length with an average play by Shakespeare. I wasn't expecting a great deal from this study as this time she would be constrained both by translation and by having to rhyme. I laid aside the first set of Psalms for the time being, as I thought I would get a better look at her most mature verse by concentrating on the latter set.

Psalms 95 - 150 yielded the following top ten scores:

1.	*3H6*	37	6.	*MAC*	24
2.	*R2*	34	7.	*KJN*	23
3.	*1H6*	31	8.	*TA*	18
4.	*2H6*	30	9.	*PER*	18
5.	*H8*	30	10.	*MND*	17

Because of my dependence on Bartlett's *Concordance* I was unable, once more, to include *The Two Noble Kinsman* or *Edward III* in the study, which most scholars believe were collaborations. Apart from these two plays, many orthodox scholars believe seven plays in the Shakespeare canon were collaborations: *Titus* and the *Henry VIs* from the 1591/92 era with several collaborators; *Pericles* from 1607 with a large input from Wilkins (Taylor 557); *Macbeth* in 1606 including later touches from Middleton (McMullen 180-99); and split authorship of *Henry VIII* by Shakespeare and Fletcher (Braunmuller 255-259). Can it be mere coincidence that all seven plays thought to exhibit the hands of writers other than Shakespeare appear in the top nine scores achieved by Pembroke's Psalms 95-150?

Pembroke had definitely completed her translation of the Psalms by 1599 and perhaps as early as 1594 (Hannay 240 f3). If orthodox scholars are correct about the chronology of his plays, then we would have to be ready to accept that Shakespeare was so familiar with the manuscript of Mary's

Antonius before 1591/2 that he adopted some of its language when writing *Titus* and the *Henry VIs*. If he didn't use her manuscript, however, these plays had to have been written before November, 1590. As per their title pages, *Titus* and *1H6* were first played by Strange's Men while the other two, first played by Pembroke's Men, were written initially for a large acting troop (Gurr 59-60).

No further trace of Pembroke's influence on Shakespeare is noted until he supposedly turned to her closet drama *Antonius* for his own *Antony and Cleopatra* in 1606 (Gurr 59-60). Middleton too was impressed enough with the vocabulary she used in the manuscript of her Psalms to have subconsciously used it in *Macbeth* in 1606 or after. In 1607, the disreputable George Wilkins scored two coups: he talked the great Shakespeare into sharing with him (of all people) the writing of a play and he got his hands on Mary's Psalms in manuscript, since their influence is found only in those sections of *Pericles* that are commonly attributed to him. Wilkins subconsciously used her vocabulary from the published Antonius as well. Her influence on the Shakespeare canon doesn't reappear until 1613 when Shakespeare decides to collaborate again and picks Fletcher, who just happens also to have read Mary's Psalms in manuscript.

All this is possible, just as it is possible to believe Southampton or William Herbert shared a manuscript written in Virginia with the playwright Shakespeare before he wrote *The Tempest*. Anything is possible. But such compounding of possibilities must lead eventually to a questioning of the orthodox authorship thesis. The problems mentioned here, being textual in nature, are apart from the numerous anomalies of the orthodox biography. Belief in the established conception of Shakespeare becomes increasingly problematic as we progress.

Pembroke and Shakespeare

How can this conjunction of all collaborative Shakespearian plays appearing in the top nine for Mary Pembroke's rare word usage in Psalms 95-150 be explained? At the very least it suggests a connection of some sort between Shakespeare and the Countess of Pembroke. I took a closer look at each play, scene by scene in relation to Pembroke's Psalms. I did the same for the other works and this time included *Edward III. 1 Henry VI* has twenty-seven scenes. Based on Psalms 95-150, Pembroke's rare words cluster at a rate of one rare word per less than twenty lines in scenes 1.1, 1.2, 1.5, 1.6, 2.2, 4.2, and 5.4. In *Antonius*, her rare words cluster in all these scenes, less 2.2, and in three others.

Is there any chance at all that these could be random spikes in the same scenes of two different works by the same author? The Shakespearian rarely used words are more prominent or cluster in greater numbers in the same scenes of *1H6* whether you use the rare words found in *Antonius* or the rare words found in Psalms II. How could this happen unless Mary was at least partially responsible for the words in *1H6*?

Nearly all orthodox scholars think *1H6* was collaboratively written. In 1985, Gary Taylor divided the writing of *1H6* into scenes written by Shakespeare, Nashe, and two other writers he named X and Y. He assigned Shakespeare scenes 2.4 and 4.2 through 4.7. He assigned Nashe most of what is now known as Act I (217). Edward Burns, in his Arden 3 edition of *1H6* agreed with

Taylor (75). From my test of other authors against *1H6* I found Marlowe, Nashe, Pembroke, Greene and Kyd had multiple scenes where their rare words clustered at a rate of one infrequently used word per less than twenty lines. Most orthodox scholars believe the process of collaboration involved the various authors dividing up the play's scenes and then going off to their studies to write their sections. Burns challenged this and I extend his interpretation (75).

For her Psalms 95-150, Pembroke scored high on *Titus Andronicus*, repeating the relatively high score from *Antonius*, and as with *Antonius*, her Psalms did not score high or form clusters of rare words in any particular scene. Therefore I don't believe that she helped write *Titus Andronicus*. Instead, I think *Titus* was a source for her language in *Antonius* and the Psalms, just as Plutarch had been a source for *Antonius*. If *Titus* was a source for Marlowe, it had to have been written by 1586.

The third supposedly collaborative effort which scores in Psalms 95-150 top nine is *2H6*. For this play Pembroke's scores in both the Psalms and *Antonius* were high overall but not particularly high in the analysis by scene. I think *2 Henry VI*, like *Titus*, influenced Pembroke's vocabulary, but I don't believe she helped write it. Gary Taylor said *2H6* was the only Shakespearian history play which could not be directly connected with the 1587 edition of Holinshed (111). I date it to 1586.

3 Henry VI is the fourth of the collaborative plays that appear in the Psalms' top nine. Unlike *Titus*, *1H6*, and *2H6*, the rare words results for *3H6* don't correlate well between *Antonius* and the Psalms. Overall scores for both were high, but for the Psalms exceptionally high with *3H6* having the greatest number of rare word correspondences with Shakespeare. In the analysis by scene, *Antonius* scored high in only two scenes while the Psalms scored high in nine. Additionally, the scenes which achieved high scores against Pembroke's two works were not the same. Something had changed since Mary Sidney completed her 1590 translation of *Antonius*.

The Countess of Pembroke c.1590

I think that, for *Antonius*, *3H6* was an influence similar to that of *2H6* and *Titus*, or of a source like Plutarch. But I believe Pembroke then edited *3H6* before the creation of the acting version by Pembroke's Men, published in octavo form in 1595. Because *3H6* was influenced by the 1587 edition of Holinshed, I think it was written that same year.

The fifth supposedly collaborative play was *Macbeth*. Most orthodox scholars follow the line that *Macbeth* was written by Shakespeare in 1606 and that Middleton later added material to scenes 1.2, 3.5, and 4.1 and perhaps touches to 1.3. (Taylor 129). Since Pembroke's clusters of rare words just happen to appear in these same scenes and those surrounding them (1.2-1.4 and 3.5-4.3), it seems highly probable that it was she who edited these scenes at some point between the play's origination and its publication in the First Folio. Some commentators think the play is too short while others feel that parts are missing (Wilson xxii-xlii). Our conclusion: it was Mary Pembroke, not Thomas Middleton, who touched up scenes in *Macbeth*, most likely to render it inoffensive to King James and his Scottish retinue.

The sixth supposedly collaborative play which also appears in the Psalms' rare word top nine is *Henry VIII or All Is True*. *H8* consists of 19 scenes including the prologue and epilogue. Most orthodox scholars believe that Shakespeare and Fletcher wrote this play collaboratively in 1613 with Shakespeare responsible for five and a half scenes and Fletcher handling the rest. Hoy, on the other hand, selected four-and-a-half scenes he thought were shared by Shakespeare and Fletcher (McMullen 448-9). To some extent, my test confirms Hoy's analysis that these four and a half scenes had mixed authorship. Regarding the 985 lines that most orthodox scholars see as Fletcher's, Pembroke shared the use of one rare word with the author every 44.8 lines, a fairly high score. In the 1164 lines orthodox scholars assign to Shakespeare, Pembroke shared one rare word with Shakespeare every 30 lines, higher still. But in the 856 lines which Hoy thought were shared by Fletcher and Shakespeare, Pembroke had one match every 22.5 lines. This result isn't as high as her scores on *1H6*, but, when her scores for the rest of *H8* are considered, it does suggest her involvement. I believe she edited these scenes as she had the scenes in *Macbeth*, probably censoring material she thought might offend either Elizabeth or James. We know that she didn't hesitate to censor her brother's *Arcadia*, she modified a number of his Psalms, and she censored *Astrophil and Stella* (Ringler 364-, 447-, 500- et seq.). As for who wrote the remaining "Fletcher" scenes, if it really was Fletcher then his vocabulary had changed radically from when he wrote *The Faithful Shepherdess*.

Pericles is the remaining collaborative play also on the Psalms' list. Mary scored high against *Pericles* for both *Antonius* and Psalms 95-150. For both, her scores for the first two acts and Gower are of similar magnitude with her scenes in *H8*, suggesting joint composition with Shakespeare. For the Psalms, Pembroke in Acts III through V of *Pericles* shared one rare word with Shakespeare every 58.3 lines. For the first two acts and Gower she shared one rare word every 24.3 lines.

Our view of Mary's possible collaboration with Shakespeare, whether a partnership, as a member of his coterie, or later as his editor, or even possibly as a combination of these, is an early attempt to explain the presence of all Shakespeare's supposedly collaborative plays in the top nine for Mary's Psalms. If this scenario derived from a scene by scene analysis is flawed, there still remains the need for an explanation of the obvious relationship between her Psalms and the Shakespeare plays that scholars have labeled collaborations.

Edward III

Although *Edward III* was not included in the Shakespeare canon at the time Bartlett compiled his concordance and was therefore excluded from my initial matrix of Shakespeare's plays, I thought it would be interesting to study as another possible collaborative play. Indeed it is interesting and presents a major problem. The scores for *Edward III* versus Shakespeare's plays dwarf the results of those plays with known authors. When the data for *Edward III* was normalized its overall average was substantially higher than even *Hamlet* for infrequently used words. Based on normalized rare words, the primary contributors to *Edward III* outside Shakespeare were Marlowe and Pembroke with Nashe, Kyd, and Greene having smaller contributions. But there was another author who contributed 110 rare words not used by any of the authors of the works studied. This

unknown author was the only writer outside of Pembroke's Psalms who used a great number of words common to both *H8* and *Macbeth*.

At present I have no way of determining who this writer might be. I didn't have data for Peele, Lodge, or Lyly to use in my test and the example for Munday was too small to test effectively. It may have been any of them. It may also have been Pembroke herself.

The top ten scores for Edward III:

1.	*1H6*	101	6.	*3H6*	39
2.	*2H6*	52	7.	*R3*	33
3.	*TA*	49	8.	*COE*	32
4.	*R2*	42	9.	*KJN*	30
5.	*LLL*	41	10.	*R&J*	29

How are we to explain such numbers in comparison with those by the other authors studied? (See Appendix, page 147.) How can we explain an additional author with rare word connections to *Henry VIII* and *Macbeth* before the death of Marlowe? One possibility for this unknown writer is that it was Mary Pembroke as she fell increasingly under the spell of Shakespeare.

Edward III was written for a smaller cast than the large companies of the late eighties and early nineties (Gurr 267), possibly a new company just formed in mid-1591. When first formed this company would have had more limited personnel than the Queen's Men, Admiral's Men, or Strange's Men. I suggest that this company was Pembroke's Men, headed by Richard Burbage. *Edward III* was a collaborative assignment for Oxford, Inc.[8] to kick-start Mary Pembroke's new toy, an acting company. What is beyond doubt, however, based on the results of this test, is that Mary Pembroke played a significant role in the story of Shakespeare.

Oxford, Inc.

If the conclusions reached above are correct, the rare word study is showing connections between Shakespeare and Pembroke. The clustering of rarely used words as shown by this scene by scene analysis is a strong indicator of some involvement in the authorship, whether collaboration or editing. If the same principles used for Pembroke are applied to the other authors studied, their rare word clusters in specific scenes suggests possible authorship for the following writers:

1. Mary Sidney Herbert, Countess of Pembroke: edited or helped write *1H6, 3H6, Pericles, Henry VIII, Macbeth, Edward III,* and probably *Richard II* and the *Henry V* choruses, and possibly the argument to *Lucrece*.

2. Thomas Nashe: collaborated on the following plays attributed to Shakespeare; *1H6, 2H6, 3H6, Titus Andronicus, The Taming of the Shrew, 1H4*; played a smaller

role on *Edward III*, and might have helped on *2H4*, *COE*, and *King John*.

3. Christopher Marlowe: (based on both of the *Tamburlaines* and *Edward II*) collaborated on *1H6*, *2H6*, *3H6*, *Titus*, *Edward III*, and possibly on *John*, *R2*, *R3*, and *COE*.

4. Thomas Kyd: based on comparison with *The Spanish Tragedy*, collaborated on *1H6*, *2H6*, *3H6*, *Titus*, and *Edward III*.

5. Robert Greene: collaborated on *1H6*, *2H6*, *3H6*, *Titus*, *Edward III*, and possibly on *The Taming of the Shrew*.

We suggest that these comparisons are indicative of the existence of a literary coterie in the late eighties and early nineties, with, at its center, a well-respected literary genius, one with access to funding. This was Shakespeare, as seen by his participation in the works cited above. The other writers in the coterie knew him as Edward de Vere, seventeenth Earl of Oxford.

"Shakespeare" in Shakespeare

Orthodox scholars have detected Shakespeare's hand in specific scenes in those plays thought to be collaborative. They don't believe other writers assisted him in these scenes. My rare words test tends to confirm this. Those scenes where none of the authors studied have rare word clusters tend to be those scenes orthodox scholars assign to Shakespeare.

In *1H6*, Shakespeare is nearly always assigned scenes 2.4, and 4.2-4.7. In scenes 2.4, 4.3, 4.4, 4.5, and 4.6, not one of the other writers studied scored high enough to suggest participation. In 4.7 (57 lines) Marlowe, Nashe and Greene scored high. In 4.2 (56 lines) Pembroke, Marlowe, Nashe and Kyd scored high.

In *2H6*, Shakespeare's scenes are not specified by orthodox scholars. Most believe Clifford's part in 5.2 is an addition at some later time by a more mature Shakespeare. The only writer studied here that scored high in this scene was Mary Pembroke, who had high scores in only two other scenes in *2H6*. In the first eight scenes of *2H6*, no writer scored high. I would think these, 5.2, and the play's plotting were Shakespeare's contribution to *2H6*.

Shakespeare's scenes are not specified by orthodox scholars for *3H6*. Scenes where no other writer had clusters of rare words are scattered throughout the play.

Although some orthodox scholars have found the vocabulary of *Titus Andronicus* suspect, others conclude that Shakespeare wrote the entire play excepting Act I. Their judgements skirt the issue of vocabulary and assign the rest of the play to Shakespeare because the play has a unified structure and an ability to plot possessed by no other author of the early nineties. But if Shakespeare's main role in this play was to plot it, then others could have written it, as the rare words test strongly suggests. Scene 3.2 is thought to have been a later interpolation by a more mature

Shakespeare and, coincidentally, this is one of a very few scenes where no other writer has clusters of rare words. Marlowe scored extremely high for Act I. Nashe scored high in scenes 2.1, 2.2, 2.4, 4.1, 4.2, 4.3, 4.4, and 5.3.

In *Edward III*, Shakespeare's scenes are thought by orthodox scholars to be primarily those involving Edward and the Countess: 1.2, 2.1, 2.2, and 4.4. For scene 1.2, only Pembroke and Kyd scored high and if the first seventeen lines (which don't, unlike the other Countess scenes, depend on Painter's novel) are thrown out, no one scores high. For scenes 2.1 and 2.2, no one achieved a high score. In scene 4.4, Pembroke, Marlowe, Kyd, and the unknown contributor scored high. High scores in those scenes not thought to be by Shakespeare were achieved many times by all writers studied. The writer achieving the highest score in each of these scenes were: 1.1 Marlowe, 3.1 Marlowe, 3.2 the unknown contributor, 3.3 the unknown contributor, 3.4 the unknown contributor, 4.1 the unknown contributor, 4.2 Marlowe, 4.3 Kyd, 4.5 Marlowe, 4.6 Marlowe, 4.7 Pembroke, and 5.1 Marlowe.

For *Macbeth*, Shakespeare's scenes are thought by orthodox scholars to be all but 1.2, 3.5, 4.1 and parts of 1.3. Of those scenes supposedly by Middleton, Pembroke scored high in all but 3.5. In scenes supposedly by Shakespeare, she scored high in 1.4, 2.1, 3.6 and 4.3. Thus her clusters are centered on and about the "Middleton" scenes. She scored high in seven of *Macbeth's* 29 scenes. No other writer scored high in *Macbeth*.

Shakespeare's scenes in *Pericles* are thought by orthodox scholars to be the final eleven scenes excluding those with Gower. Pembroke scored high in none of these scenes. Her average score in these scenes was one rare word per 58.3 lines. In those scenes thought to be by Wilkins her average score was one rare word in every 24.3 lines. Outside of Pembroke, only Nashe achieved high scores in *Pericles*. His scores were evenly distributed: high in the three scenes by Shakespeare and in the two by "Wilkins."

For *H8*, Shakespeare is thought by most orthodox scholars to have composed scenes 1.1, 1.2, 2.3, 2.4, 5.1, and the first half of 3.2. Most believe Fletcher wrote the rest. Hoy, however, thought Shakespeare split work with Fletcher in four and a half scenes beyond those he'd written alone. These split scenes are where Pembroke scored high. She didn't score high in any scene supposedly written by Shakespeare. Again, outside of Pembroke, only Nashe scored high in any scene in *H8*— one of Shakespeare's and one of "Fletcher's."

There is one other interesting play in this scene by scene analysis of the Shakespeare canon. Nashe achieved a very high score overall against *The Tempest* and had fairly high scores in several of its scenes. But he only scored very high in one scene, 4.1. A closer look found the rare words Nashe shared with Shakespeare clustered in the speeches of Iris and Ceres where the use of natural and harvesting words were pronounced as they had been in *Summer's Last Will*. Because of the close connection between every one of Nashe's works with Shakespeare I thought this might be a connection between Shakespeare and Nashe's lost play, *The Isle Of Dogs*. On the other hand, Marlowe (dead in 1593) and Pembroke also scored high in this one scene of *The Tempest*.

Conclusion

This essay leaves little room for doubt that the Countess of Pembroke played a major role in the career of Shakespeare. Her high scores in the *Antonius* comparison for those plays where Shakespeare's primary source, like hers, was Plutarch and in the three plays known to have been performed by Pembroke's Men clearly demonstrates that this rare words test measures her dependence on specific sources. Her high scores in the Psalms 95-150 comparisons indicate Shakespeare's increasing influence on her and/or her part in editing and writing certain of his plays. This might suggest that the First Folio, dedicated to her sons, may have been based on her own private collection of Shakespeare's manuscripts.

The study of the twelve other works strongly suggests the existence of a literary coterie in the late eighties and early nineties. Shakespeare's participation in this coterie and its very existence question the orthodox view of who this Shakespeare was. If the biography of the man from Stratford precludes his participation in this coterie, then he couldn't have been Shakespeare, the writer. The writing and plotting of several collaborative plays, particularly *Edward III* and *Titus Andronicus* suggest that Shakespeare was the organizing force behind this coterie. The Earl of Oxford fits this role easily. The rookie from Stratford doesn't. ———————————————————— ❦

Appendix

The top 10 results for the twelve non-Shakespearean works studied, in chronological order (positive numbers reflect the number of normalized word-sharings above average):

Christopher Marlowe
Tamburlaine I (1587)

1. 1H6   60
2. 3H6   30
3. TA   29
4. 2H6   27
5. R3   20
6. R2   18
7. R&J   18
8. KJN   18
9. COE   15
10. AYL   13

Christopher Marlowe
Tamburlaine II (1588)

1. 1H6   47
2. 3H6   40
3. TA   33
4. 2H6   29
5. MND   26
6. KJN   23
7. R2   20
8. TEM   19
9. AYL   15
10. H5   15

Robert Greene
Fr Bacon & Fr Bungay
(1589)

1. TA   22
2. 1H6   22
3. TOS   17
4. ADO   15
5. MND   9
6. 3H6   8
7. AYL   7
8. LLL   5
9. R&J   5
10. 2H6   5

Thomas Kyd
The Spanish Tragedy
(1589)

1. TA   41
2. 1H6   29
3. 3H6   26
4. R3   17
5. 2H6   13
6. TGV   13
7. TOS   10
8. R2   6
9. COE   1
10. PER   0

Robert Greene
James IV (1590)

1. 1H6   26
2. 3H6   25
3. LLL   15
4. 2H6   14
5. TGV   11
6. TOS   11
7. TA   11
8   TN   8
9. TOA   6
10. ADO   4

Anthony Munday
John a Kent (1590)

1. 1H6   0
2. COE   -4
3. TGV   -5
4. TOS   -10
5. 3H6   -11
6. TA   -12
7. PER   -12
8. LLL   -12
9. MW   -13
10. MND   -13

Mary Pembroke
Antonius (1590)

1. 1H6   37
2. PER   21
3. 2H6   19
4. TA   18
5. 3H6   11
6. JC   9
7. A&C   8
8. R2   8
9. COR   7
10. TOA   5

Christopher Marlowe
Edward II (1592)

1. TA   42
2. 3H6   31
3. 1H6   29
4. R2   21
5. 2H6   18
6. R3   10
7. TGV   2
8. JC   1
9. MAC   1
10. MND   0

Thomas Nashe
Summer's Last Will
(1592)

1. TA   68
2. 1H4   44
3. TEM   41
4. 2H6   37
5. 1H6   36
6. MW   35
7. TOS   32
8. AYL   32
9. PER   29
10. LLL   28

Mary Pembroke
Psalms 95-150 (1594?)

1. 3H6   37
2. R2   34
3. 1H6   31
4. 2H6   30
5. H8   30
6. MAC   24
7. KJN   23
8. TA   18
9. PER   18
10. MND   17

Thomas Middleton
A Trick To Catch
(1605)

1. MW   7
2. MM   4
3. TOS   1
4. AW   1
5. MND   -2
6. 1H4   -3
7. 2H6   -3
8. ADO   -6
9. COE   -6
10. MOV   -6

John Fletcher
Faithful Shepherdess
(1608)

1. MND   35
2. CYM   14
3. TGV   13
4. 3H6   12
5. AYL   10
6. PER   10
7. TA   9
8. R&J   8
9. TEM   6
10. OTH   2

Notes

1   Verbal likenesses with various writers are cited throughout the Arden editions of Shakespeare's plays. Others include G. Blakemore Evans, New Cambridge edition of *R&J*, 1984, p. 3; Jonathan Bates, Arden 3 edition of *Titus*, 1995, pp. 80-1; and Charles R. Forker, Arden 3 edition of *R2*, 2002, p. 112.

2   To replicate these charts in a short article would require far too much space. Hopefully the reader will be willing to trust the stated results.

3   Normalization makes relevant a short play like *Macbeth* when compared with a monster play like *Hamlet*. If you have a play one line long, and it contains one Shakespearian rare word, it is more significant than if a play of 3500 lines also contains only one Shakespearian rare word. Normalization here means the average line-count of all Shakespeare plays, divided by the average of the two plays under consideration, times the result, times the number of rare words shared between the two plays. Say the average number of lines of all Shakespeare plays is 2500 and play A is 2000 lines long while play B is 3500 lines long, then 2500 divided by the average of 2000 and 3500, or 2500/2750 = .91. Thus each word shared between plays A and B would be credited with .91 a word. Therefore a play of 2000 lines that registers 80 hits is more important than a play twice that length that registers the same number of hits.

4   A statistician, reading this, might comment, "Well, the total number of normalized hits could indicate that the top 4 are the ones you would expect to see at the top of all lists, because these four are the ones with the highest word to line ratios." But in my rare words study, those top 4 plays never place 1 through 4 in the top 10 standings for any plays being compared, including comparing one Shakespeare play to another. This alone shows my rare word analysis isn't generating merely random numbers.

5   Those who contribute to the discussion include editors J.C. Maxwell, Ronald Knowles, Gary Taylor, Edward Burns, John D. Cox and Eric Rasmussen (See Works Cited).

6   There is the possibility that the patron of Pembroke's Men was the Countess herself. Hannay stresses her family's "longstanding patronage of stageable drama," her uncles Leicester and Warwick's patronage of acting companies, her brother Philip's standing as godfather to the actor Richard Tarleton's child (*Phoenix* 124). The will of the actor Simon Jewell mentions funds that he expected to receive from "my ladie Pembroke" (124). That Mary's husband was the driving force behind Pembroke's Men seems questionable considering his heavy duties as President of Wales, his age, and the poor state of his health.

7   In order to support the currently accepted chronology of Shakespeare's works, the following manuscripts known to have been in the possession of the Countess of Pembroke must have been seen by Shakespeare: *The Old Arcadia*, not published until the twentieth century but used by Shakespeare in several works; The Psalms, not published until after Shakespeare's death but used in several of his works; Mouffet's *The Silkworms and Their Flies*, which wasn't published until 1599 yet was used by Shakespeare for *MND* (1595 in current chronology); and The 11th Song from *Astrophil and Stella*, which wasn't published until 1598 but was used by Shakespeare in *R&J* (1595 in current chronology).

8   Thanks to Frank Dolen, the first to use the term (in conversation on Nina Green's web-salon, Phaeton). Thanks also to Wayne Shore, who provided most of the word lists for this study.

Works Cited

Bartlett, John. *A Complete Concordance to Shakespeare*, New York: Macmillan, 1894.

Bate, Jonathan, ed. *Titus Andronicus. Arden 3. New York: Routledge, 1995.*

Braunmuller, A.R., ed. *The New Cambridge Macbeth*. New York: CUP, 1997.

Bullough, Geoffrey. *Narrative and Dramatic Sources of Shakespeare*. Vol 5. London: Routledge, 1964.

Burns, Edward, ed. *Henry the Sixth Part One*. Arden 3. London: Thompson Learning, 2000

Burrow, Colin, ed. *The Complete Sonnets and Poems of William Shakespeare*. New York: OUP, 2002.

Cox, John D. and Rasmussen, Eric, eds. *Henry the Sixth Part Three*. Arden 3. London: Thompson Learning, 2001.

Evans, G. Blakemore, ed. *The New Cambridge Romeo and Juliet*. New York: CUP, 1984.

Gurr, Andrew. *The Shakespearean Playing Companies*. New York: OUP, 1996.

Forker, Charles R., ed. *Richard II*. Arden 3. London: Thompson Learning, 2002.

Hannay, Margaret P., Kinnamon, Noel J., and Brennan, Michael G. eds. *The Collected Works of Mary Sidney Herbert*. 2 vols. New York: OUP, 1998.

_____. *Philip's Phoenix*. New York: OUP, 1990.

Hoeniger, F.D., ed. *Pericles*. Arden 2. London: Methuen, 1963.

Honigmann, E.A.J. *Shakespeare: The Lost Years*. Totawa, NJ: Barnes & Noble, 1985.

Knowles, Ronald, ed. *Henry the Sixth Part Two*. Arden 3. Walton on Thames, Surrey: Nelson, 1999.

Maxwell, J.C., ed. *Titus Andronicus*. Arden 3. London: Methuen, 1953.

McMullen, Gordon, ed. *Henry VIII*. Arden 3. London: Thompson Learning, 2000.

Melchiori, Giorgio, ed. *The New Cambridge Edward III*. New York: CUP, 1998.

Neill, Michael, ed. *The Oxford Antony and Cleopatra*. New York: OUP, 1994.

Ringler, Jr., William A. ed. *The Poems of Sir Philip Sidney*. Oxford: OUP, 1960.

Wells, Stanley and Gary Taylor, eds. *A Textual Companion*. New York: Norton, 1987.

Wilders, John, ed. *Antony and Cleopatra*. Arden 3. New York: Routledge, 1995.

Wilson, John Dover, ed. *The Cambridge Macbeth*. Cambridge: CUP, 1947.