Was Marlowe Shakespeare's Collaborator?: Computational Stylometry and the Authorship of The Three Parts of Henry VI

OTA, Kazuaki
九州大学大学院言語文化研究院

Was Marlowe Shakespeare's Collaborator?

Computational Stylometry and the Authorship of *The Three Parts of Henry VI**

Kazuaki OTA

Heminges and Condell included the three parts of *Henry VI* in the First Folio, published in 1623, which suggests that they implicitly claimed that Shakespeare authored the plays. It was not until the 18th century that editors and critics questioned this implicit claim: in 1733, Lewis Theobald doubted "whether they [these three Plays] were entirely of his Writing" (110); in 1790, Edmond Malone, echoing his doubts, suggested that "these plays were not *originally* written by Shakspeare" (429). Over the following centuries, critics and editors explored the authorship of the plays.[1] Although most scholars today believe that several playwrights, including Shakespeare, co-wrote the plays, no consensus exists as to precisely who wrote which parts. Some critics say that he wrote the whole of the plays, some reject his solo authorship and argue for the theory of multiple authorship, and others believe that Shakespeare was responsible for only small and insignificant sections of the plays. However, a couple of papers have recently appeared that claim to have solved this centuries-long issue: Hugh Craig's "The three parts of *Henry VI*" (Craig and Kinney, eds., *Shakespeare, Computers, and the Mystery of Authorship*, 40–77) and John Burrows and Hugh Craig's "The Joker in the Pack?: Marlowe, Kyd and the Co-authorship of *Henry VI, Part 3*" (Taylor and Egan, eds., *The New Oxford Shakespeare: Authorship Companion*, 194–217). Although the former does not entirely rule out the possibility of involvement by other playwrights, it argues that Shakespeare and Marlowe jointly wrote the first and second parts of *Henry VI*. Its analysis focuses on the scenes featuring Joan la Pucelle in *Part 1* and Cade in *Part 2*, concluding that Marlowe wrote most of the scenes.[2] It also suggests that *Part 3* was a collaboration, not exploring who co-authored the play. The latter contends that Shakespeare collaborated on *Part 3* with Marlowe, concluding that "there is much here to indicate that Marlowe is the author of the non-Shakespeare parts of *3 Henry VI*, and little to indicate the contrary" (217).

Craig and Burrows chiefly used two computational stylometric methods for authorship attribution in these articles, the Zeta and Delta tests (both developed by Burrows).[3] This paper examines these two forms of computational-stylistics analysis and their interpretations of stylometric results. It then argues that their interpretations of Zeta and Delta test results need to be more accurate and appropriate. Their test results provide no substantial evidence that Marlowe was Shakespeare's collaborator on the three *Henry VI* plays.[4]

1. Zeta test and its problems

There are varieties of the Zeta method. The procedure of one of the tests Craig and other contributors to

Mystery of Authorship employed when, for example, considering the possibility of Shakespeare's involvement in an anonymous text was as follows:[5]

(1) Divide each of 136 well-attributed single-author plays from the early modern period in England (27 Shakespeare plays and 109 plays by other playwrights) into 2,000-word segments. (The segment sizes range from 900 to 6,000 words, depending on the size of the target text.)

(2) Select 500 Shakespeare marker words and 500 non-Shakespeare marker words from the 136 plays. Shakespeare markers are those lexical words that often appear in Shakespeare segments but occur only rarely in segments by others. Non-Shakespeare markers are those lexical words that often appear in segments by other authors but occur only infrequently in segments by Shakespeare. (Some Zeta tests in *Authorship Companion* used 1,000 marker words instead of 500.)

(3) Count the number of Shakespeare and non-Shakespeare markers in each segment.

(4) Divide the counts of the Shakespeare and non-Shakespeare markers in each segment by the number of different words in that segment, obtaining two scores: one for the Shakespeare markers and one for the non-Shakespeare markers. Repeat the same procedure with each target segment, obtaining two scores for that segment.

(5) Plot the three groups of segments on a two-dimensional graph according to the scores of each segment.

(6) Determine whether the target text belongs to the group of Shakespeare segments or the group of segments by other playwrights based on where the three sets of segments lie.

Before examining the authorship of texts of disputed authorship, Craig and his co-researchers performed experiments to estimate the power of the Zeta method. They tried the procedure by testing plays of known authorship to see whether it could correctly assign their authorship. Figure 2.2 below shows the test result of *Coriolanus* (Craig and Kinney, "Methods" 22).[6] This test treated *Coriolanus* as if it was a play of unknown authorship. The graph shows 278 Shakespeare segments (grey diamonds), which do not include *Coriolanus*, 1,009 non-Shakespeare segments (black dots), and 13 *Coriolanus* segments (circles). Segments with a higher value for the horizontal axis and a lower value for the vertical axis are likely to be by Shakespeare, and segments with a lower value for the horizontal axis and a higher value for the vertical axis are likely to be by other playwrights. The Shakespeare and non-Shakespeare segments are separated well, if not perfectly, forming two clusters. Most of the *Coriolanus* segments fall on Shakespeare territory, confirming Shakespeare's authorship of the play and suggesting that the Zeta method is capable of authorship discrimination, though not infallible.

Figure 2.2 Lexical-words test: 2000-word Shakespeare segments versus 2000-word segments by others, with 2000-word segments of *Coriolanus*.

Figures 6.3 and 6.8 show the test results of two plays, *King John* and *Edward II*, respectively (Watt, "The authorship of *The Raigne of Edward the Third*" 127, 130).[7] The tests used the 27 plays of single-author plays by Shakespeare and 85 non-Shakespeare single-author plays from the period 1580–1619 (a
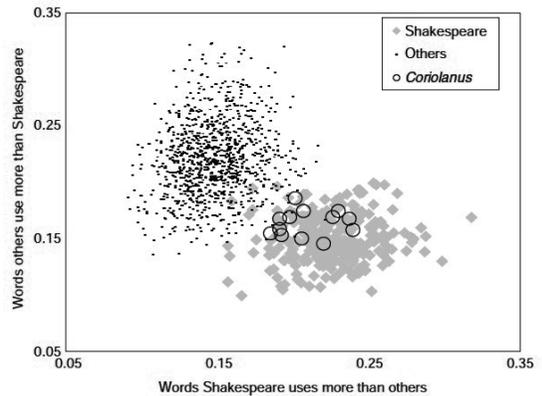
more narrowly defined "Shakespearean" than the "Early Modern" period) as control sets of texts. They divided each play into 6,000-word segments, yielding 90 Shakespeare segments and 236 non-Shakespeare segments. The graphs show the 326 segments plotted by their scores on the vertical and horizontal axes. They form two clusters with no overlap at all. The three segments of *King John* fall on Shakespeare territory, while the three segments of *Edward II*, a play by Marlowe, are all well within the non-Shakespeare cluster. These test results indicate that the Zeta method works better with 6,000-word segments than with 2,000-word segments, suggesting that the larger the segments compared, the higher the author-discriminating power of this method. Conversely, the results suggest that the smaller the segments, the lower its discriminating power.[8]
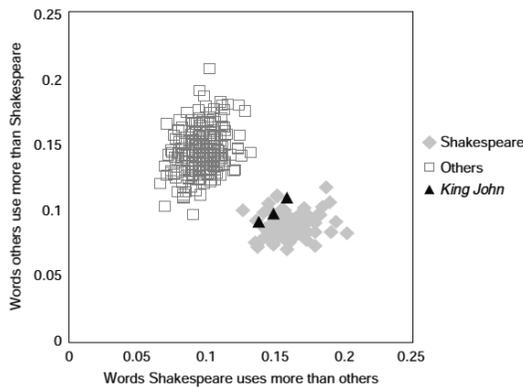

Figure 6.3 Lexical-words test: 6000-word segments of Shakespeare versus 6000-word segments of plays by others dated 1580-1619, with 6000-word segments of *King John*.
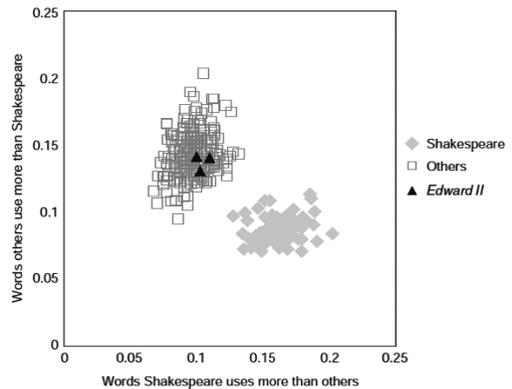
Figure 6.8 Lexical-words test: 6000-word segments of Shakespeare versus 6000-word segments of plays by others dated 1580-1619, with 6000-word segments of *Edward III*.

Figure 7.1 presents the result of the Zeta test of the "Hand-D" portion of *Sir Thomas More* (Watt, "The authorship of the Hand-D Addition" 148). The control texts for this test are the 27 single-author Shakespeare plays and the 85 single-author plays by other playwrights between 1580 and 1619, with each text divided into segments of 1,200 words. The test results form two clusters, and the Hand-D portion is well within the Shakespeare cluster, which may indicate that its vocabulary style is closer to Shakespeare than to other playwrights. However, it is difficult to determine whether Shakespeare wrote the Hand-D portion because there is considerable overlap between the Shakespeare and non-Shakespeare clusters, and we can see the Hand-D segment falling on non-Shakespeare territory, though on its fringe.

However, the extensive overlap between the Shakespeare and non-Shakespeare groups poses no fundamental problem to Watt in his test result interpretation. He and his co-researchers believe that using a bisector that divides the line joining the centroids[9] of the two clusters at a right angle can decide to which authorial group each segment belongs. Thus, they place immense importance on which side of the bisector line the two control sets of segments and the segments of the target text fall. The scatterplot shows that most non-Shakespeare segments lie above the line, while most Shakespeare segments lie below it. Watt interprets this to mean that the test correctly ascribes "the vast majority [98 per cent] of segments of known authorship" (147), attesting to the power of discrimination of the Zeta method.
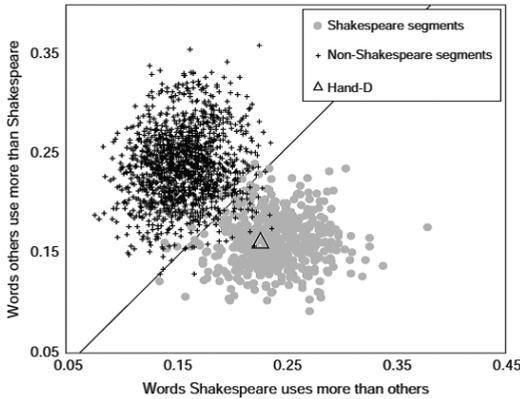
Figure 7.1 Lexical-words test: 1200-word segments from Shakespeare plays versus l200-word segments from plays by others dated 1580-1619, with the Hand-D and Addition III portions of *Sir Thomas More*.

Watt tested another six plays of known authorship to validate the method. They consist of *The Shoemaker's Holiday*, *If You Know Not Me*, *Volpone*, *The Phoenix*, *Hamlet*, and *The Duchess of Malfi*. He treated these plays as if they were anonymous and applied the test to each of them in turn. The results show that of the 105 segments of the six plays, 94 (90 per cent) fall on the "correct" side of the bisector line. With the Zeta method's power of authorship discrimination thus cross-checked, he argues that the target text (the Hand-D portion), which falls unambiguously on Shakespeare territory, is very likely to have been written by him.

However, Watt's seemingly statistically-sound argument needs more statistical robustness. He needs to conduct further tests for confirmation of its power of discrimination. First, he should test other portions of *Sir Thomas More* than the Hand-D addition and show on which side of the bisector line each of the 1,200-word segments of those portions fall. He should also test many more plays than the six because the tests of as few as six hardly prove the reliability of the Zeta method. They should include anonymous plays and plays of known authorship from the period. He should apply the test to the 1,200-word segments of these plays and present all the results for other researchers to estimate the method's reliability.

Moreover, he should perform tests using more control texts, other than the Shakespeare and non-Shakespeare sets, such as the Heywood and non-Heywood and the Munday and non-Munday sets. He should test all the 1,200-word segments of *Sir Thomas More* (including the Hand-D portion) against these sets of control texts. The results of these additional tests can help to determine whether his attribution of the Hand-D to Shakespeare is warranted.

Figure 3.9 (Craig 63) presents the result of the Zeta test Craig performed to examine the possibility of Marlowe's authorship of the "Joan sequences" in *1 Henry VI*, the sections of the play where she appears. The set of Marlowe control texts comprises six well-attributed single-author plays by Marlowe: *Doctor Faustus*, *Edward II*, *The Jew of Malta*, *The Massacre at Paris*, and *1* and *2 Tamburlaine*. Two of the three Joan segments lie below the bisector line, indicating that the two sequences use Marlowe marker words (words Marlowe uses more than other playwrights) and avoid non-Marlowe marker words (words others use more than Marlowe). This test result seems to
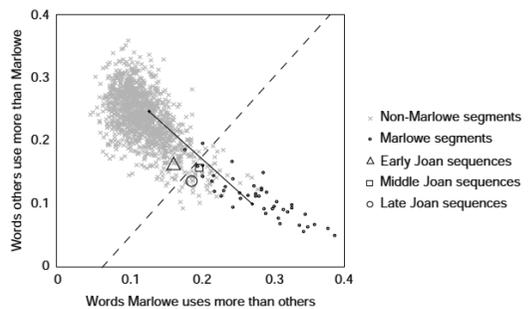

Figure 3.9 Lexical-words test: 2000-word segments by Marlowe versus 2000-word segments by others, with three sequences involving Joan de Pucelle from *1 Henry VI*.

Craig sufficient enough to suggest that the vocabulary style of the scenes involving Joan is closer to Marlowe than any other playwright. His argument, however, needs to be more convincing because the difference between the scores for the three Joan segments on the horizontal and vertical axes appears too small to be statistically significant. They are very close to the boundary between the Marlowe and non-Marlowe areas, and only a tiny variation in the scores can cause them to cross to the other side of the line. Any statistical interpretation is unreliable that draws heavily on such ambiguous results as can easily change by, for example, choice of word variables and analytic methods.[10]

A similar problem occurs in Burrows and Craig's "The Joker in the Pack," an article exploring the authorship of *3 Henry VI*, which is a sequel paper to "A Collaboration about a Collaboration: The Authorship of *King Henry VI, Part Three*" (2012), another collaborative article by them, in which they maintain they identified sixteen scenes of the play as by Shakespeare and twelve as by other playwrights. They concluded that Greene, Kyd, Marlowe, and Peele were the likeliest candidates for authorship of the non-Shakespeare scenes.[11] In "The Joker in the Pack," they examined the possibility of these four playwrights' involvement in the play by applying several forms of authorship tests.

Figure 11.5 shows the result of one of the Zeta tests they conducted to see whether Marlowe, one of the candidates, wrote what they believe to be the non-Shakespeare segments of *3 Henry VI* (Burrows and Craig, "The Joker in the Pack" 206). They tested the segments with the two sets of control texts, Marlowe's seven plays (including *Dido and Aeneas*[12]) and 56 well-attributed single-author plays by other playwrights (including 17 Shakespeare plays) from the period 1580–99. They selected 1,000 lexical words as markers of authorial difference. Figure 11.5 plotting the test result shows two clusters formed, though not well separated. The dashed diagonal line represents the bisector of the line joining the centroids of the two clusters. They added the test results of the target texts (Marlowe and non–Marlowe segments) to the scatterplot. Three of the four segments of *3 Henry VI* attributed to other writers than Shakespeare fall to the Marlowe side of the line, and five of the seven Shakespeare segments of the play lie on the non-Marlowe side. They further tested the 2,000-word segments of the other three candidates and found that no
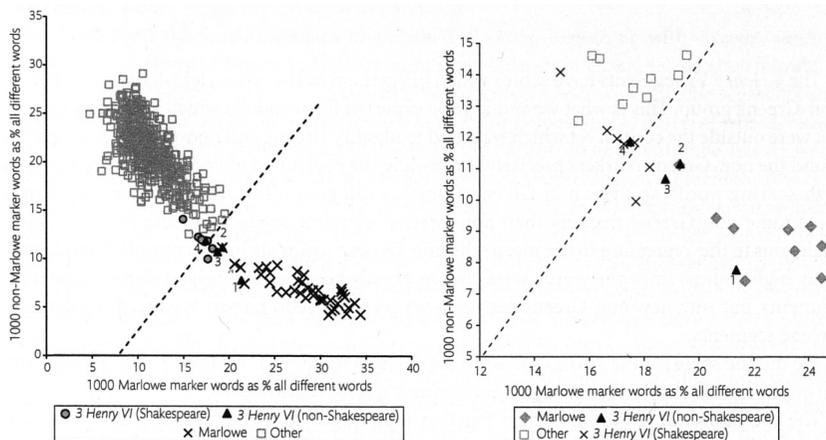

Figure 11.5 Zeta test: 2,000 word segments: Marlowe vs non-Marlowe plays, with non-Shakespeare and Shakespeare segments of *3 Henry VI*, showing the perpendicular bisector of the authorial groups.

segment of the Shakespeare or non-Shakespeare scenes of *3 Henry VI* fell on the side of any of these authors.

They also ran another variety of the Zeta test on the four authors' plays using "function-word skip bigrams"[13] instead of lexical words. The test result for the Marlowe 2,000-word segments compared to the non-Marlowe segments (Figure 11.9) shows that all four non-Shakespeare segments of the play lie on Marlowe territory. Five of the seven Shakespeare segments fall on the non-Marlowe region (Burrows and Craig, "The Joker in the Pack" 209).

The results of the Zeta tests serve as robust evidence for Burrows and Craig to contend that Marlowe is the best among the contenders for authorship of what their earlier research classified as the non-Shakespeare scenes of *3 Henry VI*. However, we can hardly agree with their ascription of the scenes to Marlowe. The reason is, first, that their interpretation of the test results needs to be revised. Indeed, the figures plotting the test results indicate that most of the Shakespeare and non-Shakespeare segments are on the "correct" side of the bisector line, and very few of the non-Shakespeare segments of *3 Henry VI* belong to the group of any other playwright than Marlowe. Nevertheless, it is not easy to interpret this as giving substantial evidence to support his authorship of the scenes. The graphs plotting the Zeta test results show that many Shakespeare and non-Shakespeare segments lie close to the boundary.[14] We cannot use these results as stable evidence for authorship discrimination since they can easily change their position and cross to the other side of the line with the slightest change of control texts and word variables. A cursory comparison of the test results presented on the scatterplots in "The Joker in the Pack" with the results of the Zeta tests with 6,000-word segments plotted on the graphs cited before (Figures 6.3, 6.8) demonstrates how ambiguous the Zeta test results of the *3 Henry VI* segments are. It is more reasonable to read these results as suggesting that the difference in vocabulary style between the Shakespeare and non-Shakespeare segments of *3 Henry VI* is not significant enough to warrant the dual authorship of the play.

Second, Burrows and Craig omitted testing its 11 segments against the 2,000-word segments of Shakespeare plays. It is hard to see why they did not carry out any Zeta test using the set of 17 Shakespeare plays for comparison with the segments of *3 Henry VI*. In their earlier study, they applied a Zeta test to the 11 segments of *3 Henry VI* to confirm the dual authorship of the play, with the result that the segments formed two clusters (Craig and Burrows, "A Collaboration about a Collaboration" 57–58). That Zeta test employed no sets of control texts for investigating to whose authorial style they are close. In "The Joker in the Pack," Burrows and Craig should have tested the 11 segments of the play using the sets of Shakespeare and non-Shakespeare plays and then compared that test result with that of the *Henry VI* segments using the Marlowe and non-Marlowe sets. Whatever conclusion they might draw from that comparative analysis, their case would have been more convincing.

Third, their selection of control texts needs to be revised. In their earlier work, Craig and Burrows used five Marlowe plays as control texts, but this study used the extended Marlowe canon, which comprised seven Marlowe plays, including *Dido and Aeneas*. They also extended the Kyd canon. While the earlier work used only *The Spanish Tragedy*, this study added two plays, *Cornelia* and *Soliman and Perseda*. Burrows and Craig wrote that they intended to "create the opportunity for a one-on-one comparison with matched canons of sufficient size to allow cross-validation" ("The Joker in the Pack" 210). However, they

should have used the same control texts in their two-part research exploring the authorship of *3 Henry VI*. Any arbitrary change in the selection of control texts, which necessarily affects the results of stylistic tests, can cast doubt on the intention of that change and may impair the credibility of statistical analysis.

Cornelia is Kyd's translation of Garnier's *Cornelie*. It is questionable whether we can establish an author's stylistic characteristics based on a text that can hardly be a work of his original authorship. This issue seems to have led them to avoid including the play regularly in their tests. Still, since the play was published under Kyd's name, they used it "occasionally" in "some broad-brush tests as a final review" (201). *Dido and Aeneas* is generally believed to be a collaborative play. Kyd's *Soliman and Perseda* is a play of disputed authorship, though attributed to him by many scholars. Any text of mixed or disputed authorship is unfit for a source of authorial markers, and they should not have used these plays as control texts for comparison. As Kinney points out, "we must assume single authorship of sections of the play to find an authorial signature" ("Authoring *Arden of Faversham*," *Mystery of Authorship* 91). Nevertheless, they used *Soliman and Perseda* "routinely" in the Zeta and Delta tests as a work of his authorship, rendering the test results less reliable.

Another area for improvement with their control texts is that they worked with canons varying in size. The Shakespeare set comprising 17 single-author plays far outnumbered those of the four candidates for authorship of *3 Henry VI*. Marlowe's canon of seven plays was the largest of the four. Peele's canon comprised four plays, Greene's four, and Kyd's three. Any corpus of three or four plays is unlikely to be a reliable basis for attribution work.

Lastly, the Zeta test results are inconsistent with some of the results of the Delta analyses Burrows and Craig performed in their earlier work and "The Joker in the Pack." (I shall discuss the Delta method employed in Craig's "The three parts of *Henry VI*" in the latter half of this paper.) The results of one of the Delta tests, presented in Table 11.3, show that *The Spanish Tragedy* is the closest play to *3 Henry VI* as a whole play (196). Table 11.5 shows another Delta analysis ranking two of the three Kyd plays among the 12 closest plays to what Craig and Burrows identified in the earlier work as the non-Shakespeare scenes of *3 Henry VI* (201).[15] Figure 11.2, which presents the average Delta distance from the non-Shakespeare scenes for the 16 playwrights in the corpus investigated, indicates that Kyd is closest to the author of these
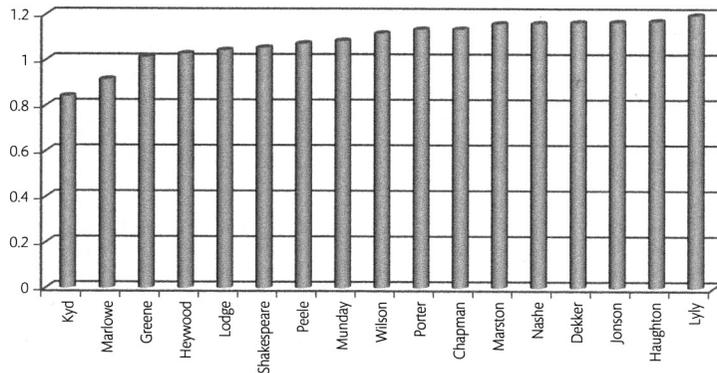

Figure 11.2 Average Delta distance from the 'non-Shakespeare' scenes of *3 Henry VI* for the 16 authors in the corpus.

scenes (203). These test results are incompatible with the results of the Zeta tests,[16] which show that Kyd's plays are less like *3 Henry VI* than the Marlowe plays. Does this tell that the Delta method is unreliable for authorship discrimination? Burrows and Craig in "The Joker in the Pack" do not seem to have much faith in the Delta method since they discarded what they "persistently" found with the Delta tests and embraced the conclusions based on the Zeta test results. If so, why did their earlier study chiefly rely on that method for classifying the scenes of *3 Henry VI*?

Furthermore, if they consider the Zeta method a powerful authorship tool, why did they not apply it to each play scene in the early work? Did they regard most scenes as too short for Zeta analysis? If so, why did they not divide the play into successive 2,000-word segments and run the Zeta test on them? Alternatively, they could have applied the Zeta test to the 108 rolling segments of 2,000 words and compared its results with those of the Delta and Iota tests they used for ascribing scenes. All this suggests that neither the Zeta nor Delta method as an authorship tool is as powerful as they make them out.

Whatever faith Burrows and Craig have in the efficacy of their methods, what is decisively wrong with their attribution work is that they adopt out of the inconsistent test results one that best serves the purpose of corroborating their theory. In "The Joker in the Pack," they based their conclusion on the results of the Zeta tests, discarding those of the Delta tests that contradicted it. However, they should not have rejected them. Instead, they should have judged that they had yet to reach a final solution to the authorship question of the play because the results of the two forms of statistical analysis failed to corroborate each other.

2. Delta test and its problems

The Delta test employed in attribution studies quantifies the distance in vocabulary style between texts.[17] Although Craig did not use the term "Delta test" or "Delta analysis" in "The three parts of *Henry VI*," he virtually employed two variants of the Delta method, carrying out distance-based authorship attribution. One method used lexical-word frequency data, and the other function-word frequency data.[18] The control texts he used consisted of the 27 well-attributed single-author Shakespeare plays and 109 well-attributed single-author plays by other playwrights, each divided into 2,000-word segments. The procedures of the tests were as follows:

Lexical-words test
(1) Apply the Zeta test to each segment of 2,000 words and find its scores for the Shakespeare and non-Shakespeare markers. Determine each segment position in two dimensions based on its scores and calculate the centroids of the two groups of Shakespeare and non-Shakespeare segments, respectively.
(2) Apply the Zeta test to each target text segment and determine its position in two dimensions based on its scores for the Shakespeare and non-Shakespeare markers.
(3) Calculate the distance of each target text segment from the centroid of each Shakespeare and non-Shakespeare cluster. Measure the difference between the two distances and express it in a single numerical value. Each segment may have a positive, negative, or zero value. The larger its positive

value, the closer to Shakespeare's style. The larger its negative value, the closer to other authors'. The value of zero means it is impossible to say to which authorial group of texts that segment is close.

Function-words test

(1) Apply the *t*-test[19] to select function words used most differently by the Shakespeare and non-Shakespeare sets of segments.

(2) Find the frequencies of the selected function words in each segment.

(3) Apply PCA (principal component analysis)[20] to the function-word data, converting the variables (function-word frequencies) for each segment into a single composite variable (principal component).

(4) Obtain the first three principal components for each segment.

(5) Determine the three-dimensional position of each segment based on the three composite variables (principal components) and identify the centroid of each of the two groups of Shakespeare and non-Shakespeare segments. (Some analyses use the first two rather than three principal components for determining the position of each segment in two dimensions.) The rest of the procedure of the function-words test follows the same process as the lexical-words test.

Craig applied these two forms of the Delta method (lexical-words and function-words tests) to six Shakespeare plays to estimate their power of authorship discrimination. He treated the plays as anonymous and examined them to see if the tests could correctly assign them. Figure 3.1 ("The three parts of *Henry VI*" 46) shows the results of the two tests. The lexical-words test classifies 50 segments above the horizontal axis as more Shakespearean than non-Shakespearean. The function-words test also classifies 50 segments that fall to the right side of the vertical axis as closer to Shakespeare. Both tests classify 43 of the 58 segments in the upper-right quadrant as Shakespearean. The "success rate" of the two methods used together is 74 per cent. Craig adds that each test, seen separately, assigns 51 of the 58 segments correctly to Shakespeare. Based on these results, he argues that the lexical-words and function-words tests, especially when used together, are powerful enough to discriminate authorial style.

Figure 3.1 Distances from Shakespeare centroids on function words and on lexical words for 2000-word segments of 6 Shakespeare plays. The procedures are as in Table 3.1, except that for each test the test play is withdrawn from the Shakespeare set.

Figure 3.2 presents the results of the lexical-words and function-words tests applied to *1 Henry VI* ("The three parts of *Henry VI*" 48). Craig maintains, "Shakespeare wrote the Temple Garden scene and the scenes depicting Talbot's last battle, but no other substantial portion" (68). The test results do not sufficiently corroborate this theory. Segment 7 (III. iv. 33–IV. ii. 56) and Segment 8 (IV. ii. 56–IV. vii. 40),[21] which
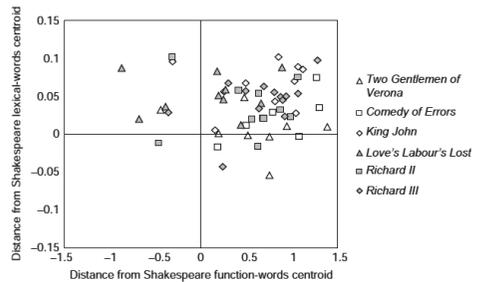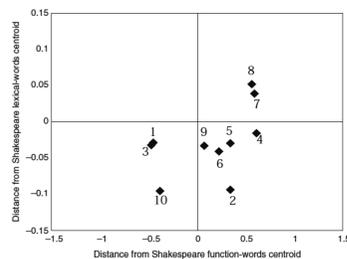
Figure 3.2 Distances from Shakespeare centroids on function words and on lexical words for 2000-word segments of *1 Henry VI*.

include scenes where Talbot appears, lie in the upper-right quadrant (segment numbers added), indicating that both tests warrant them as more Shakespearean than non-Shakespearean. However, Segment 4 (II. iii. 24–II. v. 66) includes the Temple Garden scene (II. iv), which accounts for more than half the number of lines in the segment, is below the horizontal axis, meaning that the lexical-words test classifies Segment 4 as non-Shakespearean.

Craig contends that *1 Henry VI* is a collaboration, "part by Shakespeare, part by another author or by other authors," since "the segments are more dispersed across the Shakespeare versus non-Shakespearean divides than any of the early canonical plays we tested" (47). This interpretation of the test results seems unsound. If both tests were powerful enough to discriminate Shakespeare plays and plays by other playwrights, more segments would appear in the upper-right and lower-left quadrants, and there would be fewer discrepancies between the lexical-words and function-words tests. If, as Craig claims, Shakespeare wrote no substantial portion of the play, then most segments would appear in the lower-left quadrant. Conversely, if the play were all written by Shakespeare, most segments would appear in the upper-right quadrant. Whichever the case, most segments would appear in the upper-right or lower-left quadrant. However, the function-words test classifies eight segments as "by Shakespeare," and the lexical-words test assesses eight as "not by Shakespeare." Only five of the 10 segments appear in the upper-right and lower-left quadrants. All this suggests that these forms of Delta tests (the lexical-words and function-words tests) are not as powerful in discriminating authorship as Craig claims.

Figures 3.3 and 3.4 show the results of the lexical-words and function-words tests of *2 Henry VI* and *3 Henry VI*, respectively ("The three parts of *Henry VI*" 49, 50). Thirteen of the 23 segments of the two parts of *Henry VI* (six of the 12 segments of *Part 2* and seven of the 11 segments of *Part 3*) lie in the upper-right and lower-left quadrants. Both tests warrant only three of the 12 segments of *2 Henry VI* as by Shakespeare. The lexical-words test, which classifies nine segments as non-Shakespearean, indicates that *2 Henry VI* is virtually a non-Shakespeare play. On the other hand, the function-words test, which classifies nine segments as Shakespearean, suggests that most of *Part 2* is a work of Shakespeare's authorship. Both tests place six of the 11 segments of *3 Henry VI* as Shakespearean. The lexical-words test classifies seven of the 11 segments as Shakespearean and four as non-Shakespearean. The function-words test classifies nine segments as Shakespearean, assigning most of the play to Shakespeare. Possible reasons for the inconsis-
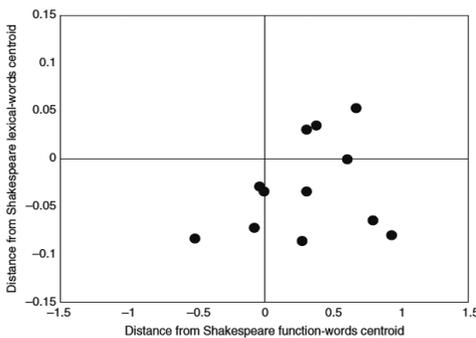

Figure 3.3 Distances from Shakespeare centroids on function words and on lexical words for 2000-word segments of *2 Henry VI*.
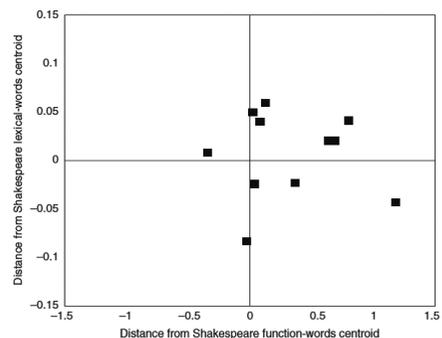

Figure 3.4 Distances from Shakespeare centroids on function words and on lexical words for 2000-word segments of *3 Henry VI*.

tent results of the two methods are:

First, the lexical-words method (the Delta procedure using the results of Zeta tests with lexical-word frequency data) can discriminate authorship, but the function-words test (the Delta procedure using the results of PCA of function-word frequency data) cannot.[22]

Second, the function-words method can discriminate authorship, but the lexical-words test cannot.

Third, neither the lexical-words nor function-words method cannot discriminate authorship. In other words, the Delta method Craig applied to the *Henry VI* plays is not a reliable authorship tool, whether using lexical-word or function-word frequency data.

Given that neither procedure produced stable results, the third reason seems the likeliest explanation for inconsistent test results. Although Craig says that the lexical-words and function-words tests are "more reliable used together than separately"(46), the mixed methods success rate of the ascription of the six Shakespeare plays (74%) is not sufficiently high to warrant their reliability. Moreover, many segments lie ambiguously—remarkably close to the horizontal or vertical axis. A slight change in the test procedure for analysis may cause them to cross to the other side of the horizontal or vertical axis, leading to different test results. Any argument based on such ambiguous test results is untenable.

The function-words test Craig applied to the three Joan sequences is also problematic. He examined whether the sequences are closer to Marlowe or any of the six playwrights who may have had a hand in *1 Henry VI*. Table 3.2 shows the test results ("The three parts of *Henry VI*" 64).[23] A positive score means

Table 3.2 Measures of closeness to Marlowe of three Joan de Pucelle sequences on function words.

	First sequence involving Joan: I.ii.22–150, I.v–II.i	Second sequence involving Joan: II.iii.1–114, III.iii, IV.vii	Third sequence involving Joan: V.ii–V.iii.44, V.iv
Greene	1.11	1.17	0.08
Heywood	0.11	1.61	0.64
Kyd	0.60	1.02	1.19
Peele	1.16	0.41	–0.34
Shakespeare	0.18	0.48	0.51
Wilson	1.12	1.72	0.78

Six PCAs were run, using word-variables with a *t*-test probability of <0.05 that the two authorial groups were from the same parent populations. The difference between the distance to the centroid for another author and the distance to the Marlowe centroid is shown. A positive score indicates the sequence is closer to the Marlowe centroid. The first two principal components were used.

that the distance between each Joan sequence and the Marlowe centroid is smaller than its distance from the non-Marlowe centroid. The scores indicate that the first and second sequences are closer to Marlowe than any of the six playwrights. The third sequence is also closer to Marlowe than any other author but Peele. Although Craig considers these scores robust evidence that Marlowe is the author of the Joan sequences, his analysis procedure has problems. First, his degree of faith in the function-words test varies depending on the conclusions to draw from the test results. In attributing the three Joan sequences to Marlowe, he drew upon the function-words test results, suggesting that he regarded this form of the test as a reliable method for determining authorship. However, in maintaining that Shakespeare wrote "no other substantial portion" of *1 Henry VI* than the Temple Garden scene and the scenes involving Talbot's

final battle, he discounted the function-words test result, which classified seven of the 10 segments of the play as Shakespearean (Figure 3.2).

Second, Craig also discounted the lexical-words test (Zeta test), which did not fully corroborate his theory. Figure 3.9 (63), presenting the test result, shows that the Marlowe and non-Marlowe segments form two clusters, though with considerable overlap. The three Joan sequences are close to the bisector line, and one falls to the side of the non-Marlowe region. This "ambiguous" lexical-words test result contradicts the "unambiguous" results of the function-words tests, which he claims support Marlowe's authorship of the sequences.

Figure 3.9 Lexical-words test: 2000-word segments by Marlowe versus 2000-word segments by others, with three sequences involving Joan de Pucelle from *1 Henry VI*.

There seems to be an inconsistency between the two Delta tests (function-words tests) applied separately to Segment 6 of *1 Henry VI* and the second Joan sequence. This segment (III. ii. 2–III. iv. 33), which accounts for two-thirds of the second Joan sequence, is classified by one of the two tests as Shakespearean. In contrast, the other test classifies the Joan sequence as closer to Marlowe than any other playwright, including Shakespeare. This discrepancy again reflects adversely on the reliability of the Delta method employing PCA of function-word frequencies.

Craig's use of PCA also lacks consistency. He used the first three principal components when testing the 2,000-word segments of the three parts of *Henry VI* to determine the three-dimensional position of each segment. In comparison, the tests of the Joan scenes used the first two principal components to calculate the two-dimensional place of each Joan sequence. The number of principal components used for analysis affects the PCA scores, and any variation in the PCA scores can lead to different conclusions about authorship attribution. He gives no explanation why the tests of the Joan sequences used the two, instead of three, principal components, which may cause doubt whether he did not choose the test procedure that would best serve his purpose.

Each of the three Joan sequences is composed of sections where Joan appears. It is not reasonable why Craig did not include I. ii. 1-21 into the first Joan sequence (I. ii. 22–150, I. v–II. 1). Act I, Scene ii constitutes a whole episode beginning with Charles, Alanson, and Reignier "marching with Drum and Soldiers." At I. ii. 21, they leave the stage and, again, "beaten back by the English with great loss," enter the scene. The same author likely wrote the whole scene, and Craig should have included the first 21 lines in the first Joan sequence. The second Joan sequence (III. ii. 1-114, III. iii, IV. vii) contains all the lines III. ii–III. iii but III. ii. 115–37. He does not explain the omission of these 23 lines. The first 32 lines of IV. vii depict Talbot's death, which Craig argues was written by Shakespeare. He gives no reason to include these Talbot lines in the Joan sequence.

The method of authorship determination used by Craig for *2 Henry VI* differs from that used for *1 Henry VI*. He employed 12 tests, a combination of three discriminant analyses,[24] six Zeta tests (lexical-words tests), and three Delta tests (PCAs with function words) to identify what parts of *2 Henry VI* Marlowe
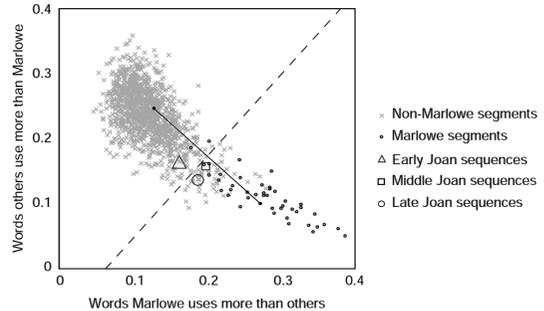
wrote. Table 3.4 shows the results of the 12 tests ("The three parts of *Henry VI*" 72). These test results confirm Craig's theory that *2 Henry VI* is a collaboration and that Marlowe, rather than Shakespeare, was the origin of the Cade scenes. This confirmation seems primarily derived from the six Zeta tests, which assign most segments to non-Marlowe. Segment 10 (IV. ii. 160–IV. vii. 121) and Segment 11 (IV. vii. 121–V. i. 13) include all the Cade scenes but IV. ii. 1–160, where he first appears. Virtually all the tests classify the two segments as Marlovian. Only one lexical-words test fails to assign Segment 10 to Marlowe. The Delta tests using PCA scores of function-word frequency data classify most segments as Marlovian. Still, although the authorship test of the Joan sequences drew on the Delta tests using function-word frequency data, Craig here discounted the function-words test results that pointed to the possibility that *2 Henry VI* was written by Marlowe rather than Shakespeare. Instead, he chose to focus on the authorship of the Cade rebellion and used the results of Zeta tests with lexical words for confirmation of Marlowe's authorship of it. This arbitrary selection of attribution methods again questions the credibility of his attribution work.[25]

Table 3.4 *Segments and sequences of* 2 Henry VI *classified as Marlowe or non-Marlowe, and Marlowe or Shakespeare, by various tests*.

Texts used	All plays	All pre-1600 plays	All pre-1600 history plays	Marlowe versus all non-Marlowe plays	Marlowe versus non-Marlowe pre-1600 plays	Marlowe versus non-Marlowe pre-1600 history plays	Marlowe versus Shakespeare plays	Marlowe versus Shakespeare pre-1600 plays	Marlowe versus Shakespeare pre-1600 history plays	Marlowe versus Shakespeare plays	Marlowe versus Shakespeare pre-1600 plays	Marlowe versus Shakespeare pre-1600 history plays
Method	Discriminant	Discriminant	Discriminant	Lexical words	Lexical words	Lexical words	PCA	PCA	PCA	Lexical words	Lexical words	Lexical words
Segment 1	✓	✓					✓	✓	✓		✓	✓
Segment 2			✓				✓					
Segment 3	✓	✓					✓	✓	✓			
Segment 4							✓	✓	✓			
Segment 5							✓	✓	✓			
Segment 6							✓					
Segment 7												
Segment 8			✓				✓	✓				
Segment 9	✓	✓	✓				✓	✓	✓			✓
Segment 10	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓
Segment 11	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Segment 12		✓	✓				✓	✓	✓			
IV.ii–x	✓	✓	✓				✓	✓	✓			✓
IV.iii–x	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Those classified as Marlowe are marked with a tick. The discriminant analysis tests use 200 function words and classify segments by author (36 authors are included in the full set, 15 in the pre-1600 set, and 3 in the pre-1600 history play set). The lexical-words tests use the 500 words that appear more regularly in one of the comparison sets and the 500 that appear more regularly in the other. The PCAs use those variables with probability <0.0005 on the *t*-test that they are from the same parent population for each text set.

In conclusion, the Zeta and Delta methods for authorship attribution are less potent than Craig, Burrows, and other practitioners of computational-stylistics studies claim. Indeed, as we have seen, some of the Zeta methods work well, especially with target texts of sufficient size (texts of 6,000 words or more). Computational stylometric analysis can quantify the difference in vocabulary use between authors or texts and visualize it through graphs and tables presenting test results. However, the lexical-words and function-words methods I have examined need revision and improvement in their procedures and interpretations of test results. They can hardly prove that Marlow was Shakespeare's collaborator on the *Henry VI* plays. It remains uncertain precisely who wrote which parts. What is certain is that their computational-stylistics test results provide no magic solution to the centuries-long issue of the authorship of the three plays.

Notes

*This article is the revised version of a paper read at the 60th Annual Conference of the Shakespeare Society of Japan, held on 1 October 2022 at Konan University.

1   For previous debates on the authorship problems of the *Henry VI* plays, see Cairncross xxviii–xxxvii, Hattaway 41–43, Craig, "The three parts of *Henry VI*" 40–43.

2   See Craig, "The three parts of *Henry VI*": "The play [*1 Henry VI*] is a collaboration. Shakespeare wrote the Temple Garden scene and the scenes depicting Talbot's last battle, but no other substantial portion. It seems very likely that Marlowe wrote the middle part of the strand of the play involving Joan of Arc" (69); "The present argument is that it was in fact Marlowe rather than Shakespeare who was the origin of these reflections of Tamburlaine in a rebel fifteenth-century clothier [Cade]" (74); "It seems that in *2 Henry VI* another powerful episode, the Cade rebellion, derives from Marlowe, while other elements, like the fraught interactions within the court after the assassination of Gloucester, can be confirmed as Shakespearean" (76–77).

3   For the Delta method, see Craig and Greatley-Hirsch, *Style, Computers, and Early Modern Drama* 44–48. For the Zeta test, see Craig and Kinney, "Glossary" 226–27.

4   Craig and Burrows are among the crucial contributors to the attribution studies that served to dictate the editorial policies of *The New Oxford Shakespeare* (2016). The *New Oxford Shakespeare* attributions, which extensively employ computational stylometric methods for the ascription of texts of disputed authorship, have drawn criticism from such scholars as Hirschfeld (2018), Rizvi (2018), Williams (2018, 2019), Rudman (2019), Vickers (2019), Auerbach (2020), and Freebury-Jones (2020, 2021). Most of these critics provide a general critical review of the *New Oxford Shakespeare* attribution methods. The present paper focuses on the problems of the Zeta and Delta methods Craig and Burrows use to explore the authorship of the three parts of *Henry VI*.

5   For this Zeta test, see Craig and Kinney, "Methods" 18–26.

6   This paper's figure and table numbers are not edited but taken from the articles examined.

7   There is an error in the caption for Figure 6.8. "6000-word segments of *Edward III*" should be "6000-word segments of *Edward II*."

8   Craig observes that "a 2000-word segment generally provides a long enough sample for reliable attribution, long enough to even out some of the local idiosyncrasies of single short passages" (The three parts of *Henry VI*" 44).

9   The centroid of a cluster of segments represents the center of the values for the cluster on each axis.

10   Rizvi doubts the efficacy of the Zeta method. In "The interpretation of Zeta test results," he argues that the current method of interpreting Zeta test results is "unsound" and that the Zeta method is "less reliable than has been supposed." He analyzes in detail the fallacy of interpreting Zeta test results using bisector lines (2–7).

11   They used the Delta and Iota methods and the Zeta test to assign scenes. The Iota test is a method of classifying "a disputed text as more likely to be by one author than another on the basis of its frequency of use of the words that appear in the first author's works and never in the second's, and vice versa" ("The Joker in the Pack" 194).

12 This play (STC 17441) refers to *Dido, Queen of Carthage*, first printed in 1594. The *Database of Early English Playbooks*, whose dates of first production Burrows and Craig follow in "The Joker in the Pack," dates its first production to 1586. The database lists no play titled *Dido and Aeneas*. It is unclear why Burrows and Craig use this title to refer to the text. They may have taken *Dido and Aeneas*, an anonymous, lost play acted by the Admiral's Men on 8 January 1598 (Harbage 88-89), for this play.

13 As they define it, a function-word skip bigram is a sequence of adjacent function words appearing together. No lexical words between them are counted. This version of the Zeta method follows the usual Zeta procedure, except that it uses "function-word skip bigrams" (instead of single words) as markers of a playwright's works for comparison with others' works. For this procedure, see Burrows and Craig, "The Joker in the Pack" 207.

14 See Figures 11.5, 11.6, 11.9, and 11.10 (Burrows and Craig, "The Joker in the Pack" 206, 208, 209, 211).

15 Marlowe's *Edward II* tops the list, *The Spanish Tragedy* stands second, and *Soliman and Perseda* third. The list does not mention *Cornelia*, and its degree of closeness to *3 Henry VI* is unknown, suggesting that its vocabulary style may be unlike the history play.

16 See Figures 11.4, 11.8, and 11.11 (Burrows and Craig, "The Joker in the Pack" 206, 209, 211).

17 John Burrows, the developer of the Delta method, calls it "Delta procedure" in "'Delta': a Measure of Stylistic Difference and a Guide to Likely Authorship," an article giving "a complete account of the procedure itself." In "The Joker and the Pack," Burrows and Craig refer to it as Delta test or analysis.

18 For the definitions of "lexical word" and "function word," see Craig and Kinney, "Glossary" 224. Unlike Burrows and Craig's Delta test in analyzing *3 Henry VI*, which uses the standardized frequencies of the 500 most common words, Craig's in "The three parts of *Henry VI* draws on the results of Zeta tests (lexical-words tests) and principal component analyses (function-words tests).

19 The *t*-test is a statistical procedure used to determine if there is a significant difference between the means of two sets of observed variables (such as word frequencies). For a brief account of the method, see Craig and Kinney, "Glossary" 225–26.

20 Principal component analysis is a data compression procedure that summarizes interrelated variables into a new composite variable (principal component). Theoretically, we can extract as many principal components as there are variables in the data. The first two or three principal components can usually suffice to represent the original data. For a detailed explanation of PCA, see Craig and Greatley-Hirsch, *Style, Computers, and Early Modern Drama* 30–39.

21 Act and scene divisions and lineation are taken from the *Riverside Shakespeare* (2nd edition).

22 Table 3 in Burrows' "Measure of Stylistic Difference" summarizes the Delta test results of 200 poems from the late 17th century, showing a steep decline in the accuracy of Delta analysis for poems with low word counts. The error rate in authorship attribution of the Delta test (using 150 word-variables) for poems of 2,001 words or more is 5 per cent, while the error rate for poems of 1–500 words is 73 per cent. See Burrows, "Measure of Stylistic Difference" 275.

23 Table 3.2 (Craig 64) contains an error in describing the second Joan sequence. It comprises III. ii.

1–114, III. iii, IV. vii rather than II. iii. 1–144, III. iii, IV. vii. Table 3.1 (51), not cited in the present paper, contains two errors in its description of Segments 9 and 10. "I. iii. 140" in Segment 9 should be V. iii. 140," and "I. ii. 140" in Segment 10 should be "V. iii. 140."

24 The discriminant analysis employed here is a statistical method that applies a mathematical procedure to function-word frequency data to determine whether each target text segment (*2 Henry VI*) belongs to the set of Marlowe segments or one of the 36 authorial groups. For discriminant analysis, see Craig and Kinney, "Glossary" 223.

25 In commenting on Burrows and Craig's arbitrary selection of Delta test results for identifying the co-authorship of *3 Henry V*, Auerbach makes a similar point: "It cannot be vaguely selectively appealed to and applied after the results have been generated, for to do so is to question the very validity of the results themselves." ("Statistical Infelicities").

Works Cited

Auerbach, David Benjamin. "Statistical Infelicities in *The New Oxford Shakespeare Authorship Companion*." *A Quarterly Journal of Short Articles, Notes and Reviews*. Published online: 21 Jan. 2019. Advance Access Article. https://www.tandfonline.com/doi/full/10.1080/0895769X.2018.1559023. Accessed 1 Mar. 2019.

Burrows, John. "'Delta': a Measure of Stylistic Difference and a Guide to Likely Authorship." *Literary and Linguistic Computing*, vol. 17, no. 3, 2002, pp. 267–87.

Burrows, John, and Hugh Craig. "The Joker in the Pack?: Marlowe, Kyd and the Co-authorship of *Henry VI, Part 3*." Taylor and Egan, *New Oxford Shakespeare: Authorship Companion*, pp. 194–217.

Cairncross, Andrew S. Introduction. William Shakespeare, *The First Part of King Henry VI*, edited by Andrew S. Cairncross. 1962. Routledge, 1994, pp. xiii–lvii.

Craig, Hugh. "The three parts of *Henry VI*." Craig and Kinney, *Shakespeare, Computers, and the Mystery of Authorship*, pp. 40–77.

Craig, Hugh, and Brett Greatley-Hirsch, eds. *Style, Computers, and Early Modern Drama: Beyond Authorship*. Cambridge UP, 2018.

Craig, Hugh, and John Burrows. "A Collaboration about a Collaboration: The Authorship of *King Henry VI, Part Three*." Deegan and McCarty, *Collaborative Research in the Digital Humanities*, pp. 27–65.

Craig, Hugh, and Arthur F. Kinney. "Glossary." Craig and Kinney, *Shakespeare, Computers, and the Mystery of Authorship*, pp. 223–27.

———, "Methods." Craig and Kinney, *Shakespeare, Computers, and the Mystery of Authorship*, pp. 15–39.

———, eds. *Shakespeare, Computers, and the Mystery of Authorship*. Cambridge UP, 2009.

Deegan, Marilyn, and Willard McCarty, eds. *Collaborative Research in the Digital Humanities*. 2012. Routledge, 2016.

Freebury-Jones, Darren. Review of *Early Shakespeare, 1588-1594*, edited by Rory Loughnane and Andrew U. Power. *Notes and Queries*, Feb. 2021, pp. 1–2.

———. "'When a man hath a familiar style': An Introduction to Authorship Studies in Early Modern Drama and Literature." *ANQ: A Quarterly Journal of Short Articles, Notes and Reviews*, vol. 33, no. 2-3, 2020,

pp. 112–21.

Malone, Edmond. "A Dissertation on *The Three Parts of King Henry VI*. Tending to shew that those Plays were not written originally by Shakspeare." *The Plays and Poems of William Shakspeare*. Vol. 6, London, 1790, pp. 381–429.

Harbage, Alfred. *Annals of English Drama, 975–1700: An Analytical Record of All Plays, Extant or Lost, Chronologically Arranged and Indexed by Authors, Titles, Dramatic Companies, &c*., revised by S. Schoenbaum, Methuen, 1964.

Hattaway, Michael. Introduction. William Shakespeare, *The First Part of King Henry VI*, edited by Michael Hattaway, Cambridge UP, 1990, pp. 1–77.

Hirschfeld, Heather. Review of *The New Oxford Shakespeare: Authorship Companion*, edited by Taylor and Egan. *Bulletin of the Comediantes*, vol. 70, no. 2, 2018, pp. 189–93.

Rizvi, Pervez. "Attribution Studies Article." Shakesper, 8 Sep. 2018, https://www.shaksper.net/archive/2018/827-september/32577-pervez-rizvi-attribution-studies-article-4. Accessed 1 Dec. 2019.

———. "The Interpretation of Zeta test results." *Digital Scholarship in the Humanities*, vol. 34, no. 2, 2018, pp. 401–18.

Rudman, Joseph. Review of *The New Oxford Shakespeare: Authorship Companion*, edited by Taylor and Egan. *Digital Scholarship in the Humanities*, vol. 34, no. 3, 2019, pp. 703–705.

Taylor, Gary, and Gabriel Egan, eds. *The New Oxford Shakespeare: Authorship Companion*. Oxford UP, 2017.

Theobald, Lewis, ed. *The Works of Shakespeare*. Vol. 4, London, 1733.

Vickers, Brian. "Valedictory." Shakesper, 19 July 2019, https://www.shaksper.net/current-postings. Accessed 20 July 2019.

Watt, Timothy Irish. "The authorship of the Hand-D Addition to *The Book of Sir Thomas More*." Craig and Kinney, *Shakespeare, Computers, and the Mystery of Authorship*, pp. 134–61.

———. "The authorship of *The Raigne of Edward the Third*." Craig and Kinney, *Shakespeare, Computers, and the Mystery of Authorship*, pp. 116–33.

Williams, William Proctor. Review of *Determining The Shakespeare Canon: Arden of Faversham and 'A Lover's Complaint'*, by Macdonald P. Jackson. *Notes and Queries*, March 2019, pp. 141–43.

———. Review of *The New Oxford Shakespeare: Authorship Companion*, edited by Taylor and Egan. *Notes and Queries*, Mar. 2018, pp. 131–34.