

## CLUSTER ANALYSIS AND THE AUTHORSHIP OF WOODSTOCK

*Thomas of Woodstock*, a play well known to students of Shakespeare but rarely performed, is the only history play in the manner of Shakespeare's *Richard II* and of Christopher Marlowe's *Edward II* that has come down to us in the form of a manuscript. It is the eighth play in a collection of fifteen manuscript plays now identified as Egerton MS. 1994 in the British Museum. Its relationship to Shakespeare's *Richard II* is such that the opening scene of Shakespeare's play requires a prior acquaintance with *Woodstock* to be fully appreciated (1). One does not otherwise sense the sting felt by King Richard in Lancaster's taunt that he is now landlord of England, not king. Nor would one suspect that, as Bolingbroke accuses Norfolk of plotting Thomas of Woodstock's death, King Richard listens, not only as judge, but as the one actually accused.

In spite of this close relationship to Shakespeare's play, authorities agree that the play *Woodstock* is the work of an anonymous writer. This play has never been claimed for Shakespeare and neither computer methods nor handwriting comparisons have ever been attempted to determine whether he is the author. Computer methods are, of course, relatively new and have yet to show a degree of success that would warrant full confidence in their use. Handwriting analysis, on the other hand, has proved sufficiently credible to establish Shakespeare's authorship for a scene in *Sir Thomas More* (2). *Woodstock* therefore provides an interesting case for authorship determination wherein the results of computer methods can be checked by a handwriting comparison.

The method of computer-aided authorship determination, cluster analysis, that was used in this investigation has been described elsewhere (3). This particular method employs objective, numerical measures of closeness between the properties of two texts that are amenable to quantification. This measure, the Euclidean distance, requires no more than simple counting and elementary arithmetic for its computation. Though the method is objective and will produce identical results when applied by different workers, prior speculation on the authorship of *Woodstock* by various authorities has only produced conflicting opinions.

The present writer makes no claim to authority in the present subject and the reader should also know that the views which prompted the present investigation are not those generally accepted. The evidence for the authorship of *Woodstock* that will be given is part of a larger investigation into the possible authorship by Christopher Marlowe of several anonymous plays that he might have written early in his career. The present author, on evidence accumulated over the past twelve years and too extensive to be listed here, has arrived at the conclusion that Christopher Marlowe was not killed in 1593 but, that after his murder staged at Deptford to escape the law, he assumed the identity of Hugh Sanford and finally died in London in 1607. The events in this latter period of Marlowe's life dispel much of the mystery surrounding Shakespeare's plays and sonnets. Source material for Marlowe's literary activity before 1593 is by comparison meager and it was with the purpose of remedying this deficiency that, after years of deliberate abstinence, the author resorted to computer aid. The temptation to do so had become irresistible for, as a twenty-three year old Marlowe burst on the literary scene in 1587, at the height of his powers with his first known composition (4), it is during these same years that we find a number

of excellent but anonymous plays that he might have written. *Woodstock* is one of these plays. If we could determine which, if any, of these plays it was that Marlowe wrote, we would be much better able to understand the development of his extraordinary genius.

The fact that this investigation is Marlowe oriented should explain why, in the thirty Elizabethan texts that were processed in the cluster analysis, there is only one accepted work by Shakespeare, *Venus and Adonis*, the "first heir of his invention". Some of the anonymous plays, however, being largely adapted or rewritten by Shakespeare, may be considered closely representative of his work as far as cluster analysis goes. Though it might to other workers, it did not seem to the author to be worth the expense of carrying on the pretence that Marlowe and the author of Shakespeare's works were necessarily two different individuals.

Wilhelmina P. Frijlinck (5) and A.P. Rossiter (6) have excellent summaries of the state of scholarship on *Woodstock*. The collection of plays, of which *Woodstock* is one, was acquired by Dulwich College from the actor William Cartwright in the middle of the 17th century (7). In 1865 the British Museum purchased the collection in the sale of Lord Charlemont's library. J.O. Halliwell, in 1870, published a limited edition of *Woodstock* as *A Tragedy of King Richard the Second, concluding with the murder of the Duke of Gloucester at Calais. A composition anterior to Shakespeare's tragedy*. A second edition by Prof. Wolfgang Keller appeared in 1879 in vol. XXV of the *Shakespeare Jahrbuch*, as *Richard II. Erster Teil. Ein Drama aus Shakespeare's Zeit*. Keller, on the assumption that the author of *Woodstock* was neither Marlowe nor Shakespeare, concluded that the unknown author of *Woodstock* borrowed more heavily from Marlowe's *Edward II* than from

Shakespeare's *Richard II*. In 1923 F.S. Boas (7) in an admiring appraisal of *Woodstock* observed :

"It has always been a *crux* to commentators on *Richard II* how its hearers or readers could be expected to be much moved by its opening scenes, of which the recent murder of Gloucester is the pivot, when the Duke himself was nothing more than a name. Again the sting in John of Gaunt's reproach to Richard for having become landlord of England instead of king is not fully comprehensible, when the strange transaction of leasing the kingdom has never been described. And even the execution of the favourites excites little interest, when we have had no concrete evidence of their misdeeds, and their most memorable utterance has been Bushy's fanciful comparison between the illusions of grief and of 'perspectives'. But all these episodes, to which Shakespeare merely alludes, are fully dealt with... in *Thomas of Woodstock* and would be deeply significant to those who knew it."

Nevertheless, F.S. Boas concluded that *Woodstock* was written after Shakespeare wrote his *Richard II*.

In 1929, Wilhelmina Frijlinck published the first accurate transcript of the *Woodstock* manuscript (5). With respect to the authorship of this play, Miss Frijlinck followed Keller in assuming that the author is neither Marlowe nor Shakespeare and in claiming that : "The author was a follower of Marlowe in his treatment and choice of plot". She also maintained, on the basis of five words inserted by a second hand into a space left blank by the scribe, that the scribe could not have been the author. In 1946, A.P. Rossiter broke with all precedent and argued that it was Marlowe who borrowed and learned from the author of *Woodstock*,

who in turn had borrowed from Shakespeare's *2 Henry VI*.

The present writer learned of the existence of the *Woodstock* manuscript in the summer of 1968 at UCLA in a course of lectures by Prof. Robert W. Dent on Shakespeare. At the time I was fully familiar with Marlowe's hand as represented in one signature and possibly in a manuscript fragment of his *Massacre at Paris* (8). Having consulted a reproduction of the *Woodstock* manuscript in Miss Frijlinck's edition, I was soon satisfied that the handwriting of *Woodstock* had certain distinct features in common with the *Massacre at Paris* fragment. Subsequently, to make a more thorough handwriting comparison of the two manuscripts I used a computer to produce word indexes of both *Woodstock* and the fragment. This computer-aided comparison produced a rather unexpected result,

The *Massacre at Paris* fragment contains a total of 311 words and a vocabulary of 179 words. *Woodstock* with a total of 25,761 words has 3,677 words of vocabulary. With a computer index to each word in both plays I did not expect any difficulty in finding handwriting matches in *Woodstock* for every one of the 179 words of vocabulary in the fragment. In the *Massacre at Paris* fragment the duc de Guise, coming upon the body of one of the king's minions, exclaims :

Thus fall Imperfett exhalatione  
wch our great sonn of fraunce Cold not effecte  
a fyery meteor in the fermament.

There are no meteors, exhalations nor a firmament in *Woodstock* and, compared to the copious astrological allusions that we find in all of Marlowe's and Shakespeare's works, but few uninspired allusions to the sun,

moon or stars. That the author of *Woodstock* knew of astrology and Ptolemaic astronomy but made little use of them is shown in two passages, first his single weak reference to astrological influence :

What heauey starr this day had dominance  
to cutt off all thy flowreing youthfull hopes.

The second is a poetic, Marlowe-like, use of Ptolemy's spheres, as the moon, personified as Cynthia, introduces a mask :

ffrom the cleere orbe of our Etheryall Sphere  
bright Cinthia comes to hunt & reuell here  
The groues of Callidon & Arden woods.

The poverty of astrological and astronomical allusions in *Woodstock* dates the composition of this play to the year 1583 or earlier, years during which a series of impressive celestial phenomena and a great earthquake, undoubtedly the topic of many sermons, were reflected in the writings of the time. The earthquake came first in April 1580, and a great comet, visible for several weeks, was observed the following October. A second comet appeared in May 1581. Finally there was an eclipse of the sun in 1582. With the much dreaded conjunction of Saturn and Jupiter to occur in 1583, it is no wonder that the study of astrological science was taken up seriously at the universities (9). By 1587 Christopher Marlowe wrote a play, *Tamburlaine*, which incorporated astrological allusion to such excess as to provoke criticism by his friends Thomas Nashe and Robert Greene. Thus, *Woodstock* must have been written some time before 1587, and perhaps as early as 1581. This early dating contradicts all previous estimates which have placed the date of composition after 1591. The later dating is based on a single passage from *Woodstock*, quoted above, which

mentions "Arden woods", on the assumption that it was borrowed from Lodge's romance, *Rosalynde*, published in 1591. Lodge, indeed, mentions a "forest of Arden", but his forest is clearly in France, probably the Forest of Ardennes, whereas the author of *Woodstock* is quite specific. His "Arden woods" is located at Plashey House in Essex"... near the Thames, circled round with trees". It is just as likely that Lodge borrowed the name "Arden" from *Woodstock* or from a map of France. Against this uncertainty we have the evidence that *Woodstock*, because of the absence of astrological allusions, must be an early play, written before 1587 and most likely about the year 1583.

This dating of *Woodstock* into the early 1580's eliminates from consideration the problem of whether the author borrowed from Marlowe's *Edward II* or from Shakespeare's *Richard II*, both written after 1591. It is also conveniently in agreement with my view that Marlowe and the author of Shakespeare's plays need not be separate individuals, a view that avoids other problems as well, such as the inexplicable influence of Marlowe on Shakespeare (10). It is hard to believe that the history plays of Shakespeare which, together with Marlowe's *Edward II* and the anonymous *Woodstock*, appear as deliberately conceived parts of a single grand design, could have been conceived and executed by three masters working independently. No other history plays by Shakespeare's contemporaries, plays by Daniel, Peele, Greene or Kydd, fall into this scheme, nor has any writer since been able to write another history play in the manner of Shakespeare and Marlowe.

The early date for the composition of *Woodstock*, an unexpected result of using a computer to find handwriting matches, will have some bearing on the handwriting comparison also, since the specimens to be compared were written nearly a decade apart. A handwriting comparison can at best establish that two hands are identical -- additional evidence is required to identify the writer. For both Christopher Marlowe and William Shakespeare, the only undisputed samples of their handwriting are their signatures to legal documents and, in the case of Shakespeare, two words "By me" from his will. We do not know that the *Massacre at Paris* fragment is in Marlowe's hand and the *Woodstock* manuscript is thought to be in the hand of a copyist, not the author. Thus, for both of these manuscripts the identity of the writer is still unresolved. Because of the substantial length of these manuscripts, compared to a signature at least, one should still be able to decide whether the two hands in question are the same or not. If, for example, the hand of the *Massacre at Paris* fragment is found to be that of the scribe of *Woodstock*, the presumption would be very strong that the hand is the hand of the author since the probable dates of composition, a decade apart, make it unlikely that the author would have employed the same scribe on these two separate occasions.

Handwriting specimens of the two manuscripts in question and from Shakespeare's will are presented in Figure 1. for comparison. A transcription of these samples is as follows (11) :



<i>line</i>	<i>Woodstock</i>	<i>Massacre at Paris</i>	<i>Shakespeare's will</i>
1	And and and	And And and	
2	B B	B	By
3	did	did	
4	Enter	Enter	
5	landlord	landlorde	
6	possessione	possessione	
7	wellcome	wellcome	ill me
8	yett	yett	

A great many more samples could be given, but those shown are typical of the degree of resemblance between the hands of *Woodstock* and the *Massacre at Paris* fragment. With respect to the differences, the *Woodstock* hand is smaller, more studied, as would be expected, since the writer compressed sixty lines to a page while the writer of the more generous hand in the *Massacre at Paris* fragment had space to spare when he was through. Though both hands employ the same forms of the letters for both English Secretary and Italian script, the *Massacre at Paris* hand is generally more cursive and economical. There are similarities in the two hands which are not shown in Figure 1., namely, that the lines of the text are not perpendicular to the vertical edge of the page, but slope upward about 2.5 degrees, and secondly, that the speech headings and some stage directions in the left margin of the manuscript slope upward about four degrees. These latter are also in a larger hand and probably in an ink of another color, now faded, entered by the writer after he had finished a page of text. All in all, the hands in *Woodstock* and in the *Massacre at Paris* fragment, though similar, can be distinguished from each other, but no more so, say, than two letters written at different times by Thomas Kydd. Given the different circumstances of composition and the interval of some ten years,

one cannot conclude that these two manuscripts were not written by the same individual.

Considering the handwriting specimens in greater detail, we find on line 1 samples of the letters "a", "n" and "d", which are identical in form compared to the 21 versions of the letter "a", the 14 versions of the letter "n", and the 14 versions of the letter "d", given by Tannenbaum (12) for English Secretary. On line 1 the *Massacre at Paris* version of the letter "A" is only slightly evolved from the *Woodstock* version, but neither of these forms is found among the thirty versions for the letter "A" given by Tannenbaum.

On the second line of Figure 1. we encounter a letter "B" that is very rare, not being found in any of the standard references (12). On the same line is a more conventional Secretary "B" in the word "By" in Shakespeare's hand. As usual, the *Massacre at Paris* version of this letter is more rounded and generous. We note also that the two letters "B" from *Woodstock*, though similar in appearance, are formed with different strokes as if the writer had in mind the form of the letter and was still experimenting with various ways to produce it. The first *Woodstock* "B" is formed with the same pen strokes used to form the *Massacre at Paris* "B", but the second *Woodstock* "B" actually bears a closer resemblance to the latter. This capital letter "B" is the strongest evidence that the *Woodstock* and *Massacre at Paris* manuscripts were written by the same writer.

On line 4 the writer of both manuscripts changes to the Italian script for stage directions, in this case for the word "Enter". In both examples, the style of each of the five letters is the same, remarkably so even for words written ten years apart by the same writer. What appears to be another

unique point of similarity on line 5, namely, the manner of joining the two letters "dl" in the word "landlord", is, in fact, very common. The fact that the words are spelled respectively "landlord," and "Landlorde" is also without significance, such variations, in even a single text, being rather common in Elizabethan times. On line 6 the two words "possessione" again appear very much alike except for size, and both employ for the letter "p" a character which is easily confused with the English Secretary letter "x". According to Miss Frijlinck this use of an "x" for a "p" is a habit with the scribe of *Woodstock*, though the one shown in the word "possessione" is still listed by Tannenbaum as a "p". Certainly, the version of the letter "p" used in the *Massacre at Paris* fragment has degenerated even further into the Secretary letter "x".

The word "wellcome" on line 7 shows an exception to the general rule that the hand in the *Massacre at Paris* fragment is more cursive than that of *Woodstock*, this word being formed with three breaks for *Woodstock* and four for the fragment. Elsewhere in the *Woodstock* manuscript the word welcome is spelled "wellcom", and there it is formed with only two breaks. The handwriting samples "ill" and "me" on line 7 are excerpted from Shakespeare's signature as an example of another hand using letters appearing in the word "wellcome."

On the last line the words "yett," spelled the same and looking very much alike illustrate what has already been observed. The *Massacre at Paris* hand is more generous, rounded and less studied than the *Woodstock* version. Simpler strokes are used to form the letter "y" in the *Massacre at Paris* example. Shakespeare's "y" is shown in line two for comparison.

he could think of nothing better, would have been penned in with manuscript in hand. Throughout the play the writer strove for balance and symmetry and, according to F.S. Boas (7) "... historical accuracy is ruthlessly sacrificed to secure a symmetrical balance of scene and character". Thus the author's difficulty with line 331 began some lines earlier :

Tris : Interrupt me not, those dayes thow knewst I say  
from whence I did become a plodding clarke  
from wch I bounst as thou dost now in buckram  
to be a pleading lawyer (& ther I stayd,)  
till by the king I was chiefe lustice mayd.

The balance between plodding clarke and pleading lawyer is defective in that the latter does not fit the meter. The author needed a one syllable synonym for lawyer and not being able to find it settled for a less satisfactory substitute.

There is, moreover, persuasive internal evidence that the *Woodstock* manuscript is in the author's hand. It is true, as Miss Frijlinck says, that the manuscript was copied from a rough draft, but not of a finished play. In the course of writing the smooth copy, the author continued to make changes, additions and deletions in his last opportunity to fuse the characters and scenes, that he had separately conceived and set down on odd scraps of paper, into a coherent, dramatically effective vehicle for the stage. Fortunately for us, the writer was economical of paper, evidenced by his cramming sixty lines to a page, and he never discarded and rewrote a page of his smooth copy if it had extensive deletions. It is in these deletions and their replacements, preserved for posterity, that the hand of the author is clearly evident. If, in a moment of inattention, he has allowed an actor's speech to wander off and detract from the effectiveness of the play, he will cross out some 20 lines and start anew and back on key. Revisions of this type can only be made by an author who is thoroughly

In summary, as to handwriting evidence, the unique letter "B" together with the consistent use of the same form of both English Secretary and Italian letters support the conclusion that both the *Woodstock* and the *Massacre at Paris* manuscripts were written by the same individual. Any differences in the two hands are of a kind that one would expect, given the different times and circumstances of composition. The more fluent hand in the *Massacre at Paris* fragment could easily have evolved from the hand in *Woodstock*, described by Miss Frijlinck as ". . . written in a regular and well-formed English hand with a good deal of individual character, and clearly of a literary rather than a professional type."

Pursuing the conclusion that the *Woodstock* manuscript and the *Massacre at Paris* fragment are in the same hand we can no longer accept Miss Frijlinck's view that: "There can be little doubt that the work is that of a scribe, copying from a rough draft, and not of the author himself. This appears from the nature of the corrections . . . At one point (line 331) the first half of a line was left blank, apparently because the scribe could not read the original, and subsequently supplied by another hand." Would this second hand then be by the author? Miss Frijlinck ventures no opinion. Examination of the passage in question does raise some doubt to Miss Frijlinck's generally accepted view. First, the two halves of line 331 :

to be a pleading lawyer (& ther I stayd,)

are not clearly in different hands as any apparent difference is easily explained by the fact that the words "(& ther I stayd)" were written, as Miss Frijlinck would agree, while the writer was still at his standish, settled in a comfortable writing position, whereas the addition, "to be a pleading lawyer, " perhaps supplied at the last minute by the author himself when

familiar with the structure of the play and the development of the plot as a natural consequence of the interplay and motivations of the characters. Thus, contrary to previous opinion, the deleted passages and their replacements strongly favor the conclusion that the author himself was the scribe of *Woodstock*. How could he afford a scribe if he could not even afford the paper to replace spoiled sheets in his smooth copy ? Thus also, no copy of *Woodstock* was ever made, and certainly, the author did not keep one for himself, for, if he is also the author of *Richard II*, in writing the latter play he forgot which characters he had already killed off--an amusing situation should anyone attempt to perform *Richard II*, parts 1 and 2, in succession and bring once-dead courtiers back to life.

The relationship of the *Massacre at Paris* fragment to the printed play *Massacre at Paris* also appears to be that of a rough draft to a smooth copy. The fragment, written in Marlowe's flamboyant style, is severely pruned and edited for the play, though with no loss of essential detail. The fragment has no *raison d'être* except as a rough draft of one of the many brief kaleidoscopic scenes of which Marlowe's play is composed. These scenes, transcribed in great haste it would seem, were combined into the smooth copy of the play that was delivered to Henslowe in 1592. Though it may be argued that a scribe could transcribe a smooth copy from a rough draft by the author, one could hardly attribute the writing of the original rough drafts to a scribe. Thus, on this basis alone, the *Massacre at Paris* fragment must be the first draft of a scene in the hand of the author himself, Christopher Marlowe.

That a single hand, the hand of the author in both cases, produced the *Woodstock* and *Massacre at Paris* manuscripts is confirmed by the spelling

of words in *Woodstock* and in Marlowe's printed works. In the *Woodstock* manuscript we find spellings such as mee for me, yee for ye, bee for be, foorth for forth, toomb for tomb. In Marlowe's works, such as *Hero and Leander* and *Edward II*, in the face of any house style or compositor preference that might have prevailed, we find similar archaic spellings, namely : wee, shee, bee, hee, toong (for tongue), foorth, hoong (for hung) and woon (for won). In both *Woodstock* and in Marlowe's works there is no consistency in the use of these spellings.

Computerized cluster analysis was not employed in the handwriting comparisons, first, because of a lack of handwriting specimens for most of the texts in question, and secondly, because, as yet, no convenient method has been developed for quantifying handwriting characteristics. Cluster analysis was therefore restricted to various frequency distributions, some well-known, some of them new, readily produced by the computer from machine-readable text. To determine the identity of a group of unknown objects, a cluster analysis must not only include the objects in question, but also objects from two other groups : those to which the unknowns are suspected to belong, and those to which the unknowns are known not to belong. To determine the authorship of just one play, like *Woodstock*, using cluster analysis requires the availability of texts in machine-readable form to a total of hundreds of thousands of words. Fortunately, when a number of contemporary texts must be clustered to ascertain authorship, the size of the required corpus is not much increased. This was the case with respect to *Woodstock* which was only one of about a half-dozen anonymous plays that were processed in the quest for early work by Marlowe.

A total of thirty texts, not all of them distinct, some 450,000 words in all, were used in the cluster analysis. The works of Marlowe were keypun-

ched from photostat copies of the best Elizabethan editions available, carefully proofread with the aid of a computer, and converted to modern American spelling. Spelling variants were not used as a possible criterion of authorship. The balance of the thirty texts have not been proofread as thoroughly and reduced to a standard spelling but, as will be shown, cluster analysis, because it deals with masses of data, is not much influenced by a small percentage of errors. The following list is a summary description of the Elizabethan texts used in the cluster analysis. Each item is headed by the four letter acronym used to identify the text in computer processing.

TAM1, the first part of *Tamburlaine*, by Christopher Marlowe; 17748 words, 3287 words of vocabulary, average word length : 4.416 letters, average sentence length : 18.584 words.

TAM2, the second part of *Tamburlaine*, by Christopher Marlowe; 18122 words, 3287 words of vocabulary, average word length : 4.371 letters, average sentence length : 20.711 words.

DF16, *Doctor Faustus*, the 1616 edition, by Christopher Marlowe; 16747 words, 3080 words of vocabulary, average word length 4.247 letters, average sentence length : 11.331 words.

JEWM, *The Jew of Malta*, by Christopher Marlowe; 18309 words, 3040 words of vocabulary, average word length : 4.056 letters, average sentence length : 12.690 words.

DIDO, *Dido, Queen of Carthage*, by Christopher Marlowe; 13707 words, 2803 words of vocabulary, average word length : 4.143 letters, average



sentence length : 17.920 words.

MAPA, *The Massacre at Paris*, by Christopher Marlowe; 10351 words, 1944 words of vocabulary, average word length : 4.108 letters, average sentence length : 14.141 words.

EDW2, *Edward the Second*, by Christopher Marlowe; 21099 words, 3191 words of vocabulary, average word length : 4.140 letters, average sentence length : 12.514 words.

H&LM, the first two sestiams of *Hero and Leander*, by Christopher Marlowe; 6313 words, 1967 words of vocabulary, average word length : 4.380 letters, average sentence length : 21.769 words.

LUCN, translation of the first book of Lucan's *Pharsalia*, by Christopher Marlowe; 5499 words, 1981 words of vocabulary, average word length : 4.563 letters, average sentence length : 40.434 words.

OVID, *Ovid's Elegies*, a translation of Ovid's *Amores* by Christopher Marlowe; 20242 words, 4227 words of vocabulary, average word length : 4.267 letters, average sentence length : 15.631 words.

H&LC, the last four sestiams of *Hero and Leander*, by George Chapman; 14468 words, 3297 words of vocabulary average word length 4.362 letters, average sentence length : 36.908 words. This great disparity in sentence length, a characteristic of Chapman, compared to the sentence length for Marlowe's first two sestiams (above) should forever dispel any doubt that Chapman, not Marlowe, is the author.

V&A1, approximately the first half of *Venus and Adonis*, by William Shakespeare; 3573 words, 1266 words of vocabulary, average word length 4.252 letters, average sentence length : 26.080 words.

V&AD, the whole of *Venus and Adonis*, by William Shakespeare; 7283 words, 2164 words of vocabulary, average word length : 4.283 letters, average sentence length : 25.826 words.

WOOD, *Thomas of Woodstock*, anonymous; 25761 words, 3745 words of vocabulary, average word length : 4,260 words, average sentence length : 13.150 words.

SPTR, *The Spanish Tragedy*, by Thomas Kydd; 21383 words, 3693 words of vocabulary, average word length : 4.278 letters, average sentence length : 15.048 words,

LEIR, *The Tragical History of King Leir*, 1605, anonymous; 21709 words, 3098 words of vocabulary, average word length : 4.067 letters, average sentence length : 19.263 words.

YRK1, *The Contention of York and Lancaster*, Part 1, anonymous; 17637 words, 2727 words of vocabulary, average word length 4.097 letters, average sentence length : 14.921 words. Since this play is largely assimilated in Shakespeare's 2H6 (the second part of *Henry VI*) it may be considered representative of the latter for clustering purposes.

YRK2, *The Contention of York and Lancaster*, Part 2, anonymous; 17803 words, 2873 words of vocabulary, average word length : 4.146 letters, average sentence length : 15.190 words. For reasons given above, this play

may be considered representative of 3H6 for clustering purposes. This particular text was converted to modern American spelling.

YORK, the same as YRK2, except that the spelling is largely English; 17857 words, 2861 words of vocabulary, average word length : 4.137 letters, average sentence length : 15.172 words.

TOAS, *Taming of A Shrew*, anonymous; 12752 words, 2003 words of vocabulary, average word length : 3.930 letters, average sentence length : 14.985 words. This play was completely rewritten by Shakespeare as *The Taming of The Shrew*.

ARDN, *Arden of Faversham*, anonymous; 20197 words, 3023 words of vocabulary, average word length : 4.018 letters, average sentence length : 13.777 words.

FBFB, *Friar Bacon and Friar Bungay*, by Robert Greene; 16819 words, 3084 words of vocabulary, average word length : 4.208 letters, average sentence length : 15.290 words.

KJN1, *The Troublesome Reign of King John*, part 1; anonymous, 14494 words, 2909 words of vocabulary, average word length 4.214 letters, average sentence length : 18.823 words.

KNJ2, *The Troublesome Reign of King John*, part 2, anonymous; 9903 words, 2207 words of vocabulary, average word length : 4.179 letters, average sentence length : 18.685 words.

KJ12, the texts of KNJ1 and KNJ2 combined, anonymous; 24386 words,

3887 words of vocabulary, average word length : 4.200 letters, average sentence length : 20.492 words. Except for some rewriting and bowdlerization, this combined play is practically identical to Shakespeare's *King John*.

WILL, *Will Summers' Last Will and Testament*, by Thomas Nashe; 16889 words, 3980 words of vocabulary, average word length : 4.172 letters, average sentence length : 27.240 words. This play demonstrates what many may have suspected, that Nashe in his writings exhibits the largest vocabulary of any Elizabethan.

FVH5, *The Famous Victories of Henry the Fifth*, anonymous; 13035 words, 1652 words of vocabulary, average word length : 3.903 letters, average sentence length : 11.198 words.

CORN, about a third of *Cornelia*, translated from the French of Garnier by Thomas Kydd; 6434 words, 1936 words of vocabulary, average word length : 4.436 letters, average sentence length : 23.061 words.

PEMB, a letter to Queen Elizabeth from the Earl of Pembroke, 1592; 2790 words, 704 of vocabulary; average word length : 4.572 letters, average sentence length : 45.738 words. This letter exhibits the small vocabulary and long sentences typical of legal and administrative documents.

RIC3, *The True Tragedy of Richard III*, anonymous; 19949 words, 2800 words of vocabulary, average word length : 4.094 letters, average sentence length : 18.185 words. This play, bearing but little resemblance to Shakespeare's *Richard III*, may be an early composition by Christopher Marlowe.

The above texts were processed by a computer program, CONSTAT (3), which produced for each text eight percentage frequency distribution curves, each curve limited to forty points. A second computer program was then used to arrange similar curves into clusters, using the Euclidean distance between two curves as the measure of closeness. The clustering computer program also provided a table showing the Euclidean distance between any two texts, 435 distances in all for the thirty texts, these distances being computed for all eight percentage frequency distribution curves. Each text is then taken as the nucleus of a cluster bearing its name, loose clusters being a group of texts that are less than the average distance from the nucleus, and tight clusters being the smaller group that are less than the average minus one standard deviation from the nucleus. Texts which are distant from all other texts by more than the average distance and for which not even a loose cluster can be found are listed as mavericks, such texts typically being radically different in theme or genre from other members in the group. The computer program also provides a list of the text pairs, 435 in the present case, ranked according to distance, the two closest texts being listed first.

The eight frequency distributions that were used in the cluster analysis are the following :

1. Word length distribution or the relative percent frequency of occurrence of one-letter, two-letter and up to 40-letter words. This word length frequency distribution, introduced in 1887 by Dr. Thomas Corwin Mendenhall, President of the American Association for the Advancement of Science, has proved to be one of the more reliable discriminators for authorship.
2. Sentence length distribution, or the relative frequency of one-word, two-word, and up to 40-word sentences. Since there are far fewer sentences than

words in any text, this frequency distribution is rarely smooth and will even display periodicities for texts in verse.

3. The relative percent frequency of occurrence of the 26 letters of the alphabet and the ten Arabic numerals. This distribution (13) was adopted to take advantage of the larger sample size that these characters provide.

4. The relative percent frequency distribution of the forty most common words in Elizabethan English, namely, the words : and, the, to, I, my, of, you, a, that, in, for, me, with, be, not, but, his, your, is, this, it, will, have, thou, he, as, so, him, all, what, shall, thy, no, are, now, by, do, thee, if, then.

5. The relative percent frequency distribution of forty prepositions, in effect, all the prepositions used in Elizabethan English, namely : to, of, in, for, with, by, on, from, at, upon, unto, up, out, before, down, about, ere, against, within, over, after, near, without, through, amongst, next, afore, betwixt, between, under, above, except, behind, beside, amidst, along, per, among, below, outside.

6. The relative percent frequency distribution of "-tives", that is, of connectives, negatives, interrogatives and demonstratives, namely, of the words : and, but, as, so, or, if, hence, therefore, though, unless, since, lest, because, either, not, no, never, nor, nay, cannot, nothing, none, neither, naught, what, then, why, how, when, where, who, there, hence, which, whence. Because an author must use words in this group but generally has some choice and may prefer one word over another, say, to express a negative, this recently adopted word list shows promise as an author discriminant.

7. The relative percent frequency of words of vocabulary that are used once,

used twice, and up to forty times in a text, the percentage being with respect to the number of words in the text vocabulary. The use of this distribution, which I call the Yule distribution, is due to the actuarial statistician George Yule (14), who regarded words of vocabulary as he had once regarded accident-prone people; for example, a word occurring once is less accident prone than one occurring more often in a text. Paradoxically, Yule could not find the number of accident-free members of this population, that is, the number of words that might have occurred but which did not appear at all in a given text. He therefore devised his famous characteristic K, which he thought would be a constant for a given author. In this Yule was mistaken and his characteristic K has the further disadvantage that it is ambiguous. that is, the same value of K may be obtained with widely different Yule distributions. The Yule distribution, as used here, though more reliable than the characteristic K, has proved to be a rather weak authorship discriminant.

8. Another form of Yule's distribution, namely, the frequency of words used once times one, of words used twice times two, up to words used forty times times forty. This percentage distribution is with respect to the total number of words in the given text.

Though the results of the cluster analysis provide information on the authorship of all the anonymous texts that were processed, only the results pertaining to the authorship of *Woodstock* are presented here, an exception being made when *Woodstock* clusters closely with another anonymous text. In the latter case data relative to the authorship of the second anonymous text is also presented. First, with respect to the sensitivity of cluster analysis to corrupt or slightly altered texts, the two versions of *The Contention of York and Lancaster*, Part 2, one being in English and the

other in American spelling, differed in the eight distribution curves just described by the following percentages : 0.245, 0.997, 0.039, 0.262, 0.043, 0.233, 0.494 and 0.743, being in all cases the smallest percentage Euclidean distances obtained. As the ratio of the above minimum distances to the corresponding mean distance was always less than 11 %, the error in Euclidean distance resulting from the use of a slightly corrupt text is not significant.

Following is a summary of the results of the cluster analysis with respect to *Woodstock*. Values of the mean and standard deviation are given for the 435 Euclidean distances between the thirty texts processed, all numerical values being expressed in percent.

1. Word length. Mean : 4.85, standard deviation : 2.805, minimum : 0.379 between KJN1 and KJN2, maximum : 14.890. The seven texts closest to *Woodstock* and their distances are KJ12 1.34, KJN2 1.38, KJN1 1.45, DF16 1.56, FBFB 1.61, YRK2 1.94 and MAPA 1.96. Since the closest text to the anonymous *King John* is also *Woodstock*, the author of the one is likely the author of the other. *King John* appears to be the third in a series of plays in which the author, in the prologue, mentions an earlier play, presumably a play of his own. Thus the prologue to *King John* mentions an earlier *Tamburlaine*, which in turn has a prologue that refers to an earlier, unnamed comedy. Since Marlowe is the author of *Tamburlaine*, he is then probably the author of *King John*, and the earlier comedy might have been *The Jew of Malta*.

2. Sentence length. Mean : 11.76, standard deviation : 4.900, minimum : 3.154 between KJN1 and KJN2, maximum : 26.325. The seven texts clo-



sest to *Woodstock* and their distances are : DF16 4.62, FVH5 5.81, FBFB 5.92, MAPA 6.38, YRK1 6.52, SPTR 6.54 and JEWM 7.02.

3. Alphabet frequency, Mean 2.07, standard deviation : 0.690, minimum : 0.304 between KNJ1 and KNJ2, maximum 4.141. The seven texts closest to *Woodstock* and their distances are : EDW2 1.22, FBFB 1.43, TAM1 1.47, MAPA 1.59, TAM 1.65, LEIR 1.69 and KJN2 1.75.

4. Common words. Mean : 7.81, standard deviation : 3.340, minimum : 1.295 between KJN1 and KJN2, maximum 17.279. The seven texts closest to *Woodstock* and their distances are : EDW2 3.76, YORK 3.82, YRK2 3.88, YRK1 4.15, MAPA 4.34, DF16 4.42 and FBFB 4.46.

5. Prepositions. Mean : 9.24, standard deviation : 3.630, minimum : 1.382 between KJN1 and KNJ2, maximum 21.169. The seven texts closest to *Woodstock* and their distances are : YORK 3.34, YRK2 3.35, SPTR 4.44, JEWM 5.32, RIC3 5.45, EDW2 5.51 and ARDN 5.57. Since the closest text to the anonymous YORK is also *Woodstock*, the author of *Woodstock* is likely the author of this text, namely YORK. The text next closest to YORK is SPTR, *The Spanish Tragedy* by Thomas Kydd at a distance of 3.34. Though works by Thomas Kydd bear little resemblance in style to the works of Marlowe, Kydd's *Spanish Tragedy* clusters fairly closely with several works of Marlowe, *Edward II* and *Doctor Faustus* in particular. This may be due to the fact, as both Nashe and Kydd himself affirm, that Kydd roomed with Marlowe about the year 1589 and that Kydd at this time was employed as a scrivener or scribe, possibly turning out smooth copies of Marlowe's plays, and as a translator, translating *Cornelia* from the French of Garnier and, as Nashe seems to imply, the story of Hamlet from the *Histoires Tragiques* of Belleforest. It would appear that Kydd was a junior partner or apprentice

working closely with Marlowe and to some extent influenced by him.

6. Connectives, negatives, interrogatives, demonstratives. Mean : 9.88, standard deviation : 4.25, minimum : 1.543 between KJN1 and KJN2, maximum : 24.413. The seven closest texts to *Woodstock* and their distances are : EDW2 4.78, YRK1 4.86, TOAS 5.05, YORK 5.12, YRK2 5.15, DF16 5.42 and SPTR 5.55.

7. Yule's distribution. Mean : 6.40, standard deviation : 4.369, minimum : 0.933 between SPTR and DF16, maximum : 21.002. The seven closest texts to *Woodstock* and their distances are : RIC3 1.33, LEIR 2.97, TAM2 3.17, TOAS 3.34, YRK1 4.06, TAM1 4.48, FVH5 4.61. That this distribution is somewhat unpredictable is evidenced by the fact that in the first six distributions just given, the two closest texts were always KJN1 and KJN2, namely, the two parts of the anonymous *Troublesome Reign of King John*, whereas for Yule's distribution, the two closest texts are two plays of clearly different style and theme, plays by two known authors, namely *Doctor Faustus* by Christopher Marlowe and *The Spanish Tragedy* by Thomas Kydd.

8. Yule's modified distribution. Mean : 6.98, standard deviation : 4.571, minimum : 1.676 between ARDN and EDW2, maximum : 20.149. The seven closest texts to *Woodstock* and their distances are : TOAS 2.26, RIC3 2.30, LEIR 2.46, YRK1 2.52, KJ12 2.53, EDW2 2.56 and ARDN 2.82.

To summarize the results of the cluster analysis with respect to *Woodstock*, for four out of the eight frequency distributions a play by Christopher Marlowe, *Doctor Faustus* in one case and *Edward II* in the other three cases,

were closer to *Woodstock* than any of the other texts processed. In two instances anonymous plays, namely *The Troublesome Reign of King John* and *The Contention of York and Lancaster*, part 2, were closer to *Woodstock* than any other. These plays are assimilated with some rewriting in Shakespeare's *King John* and *Henry VI*, part 3, and, if originally by Shakespeare himself, would argue for Shakespeare's authorship of *Woodstock*. The clusters obtained using Yule's distributions differed considerably from the clusters produced by the first six distributions. For all eight frequency distributions, however, one or more known works of Marlowe are found in a tight cluster with *Woodstock* as nucleus. If we assign the *Troublesome Reign of King John* and the *Contention of York and Lancaster* to Marlowe, he is by all odds the author of *Woodstock*. But even if one assigns these plays to another author, the preponderance of evidence would still favor Marlowe's authorship.

Handwriting comparisons and cluster analysis are in satisfactory agreement as to the authorship of *Woodstock* and support the conclusion that the author is Christopher Marlowe. Marlowe's stamp is acknowledged by earlier authorities such as Keller, Frijlinck and Rossiter, who only disagree as to whether Marlowe was the borrower or the giver. If, in ascribing the authorship of *Woodstock* to Marlowe, we have correctly fitted another piece of the Elizabethan dramatic authorship puzzle, then we should expect to see other pieces of the puzzle fall into place. This predictive property of a scientific conclusion has already been demonstrated for Marlowe's authorship of *Woodstock*.

In a recent letter, Dr. William Urry, the authority on Marlowe's family in

Canterbury, has urged the close contracts between "Lord" Cheney, warden of the Cinque Ports, and the Marlowes in Canterbury as evidence that only Christopher Marlowe could have written *Arden of Faversham* (15). In the play *Woodstock*, a non-historical Sir Thomas Cheney appears prominently in the opening scene and, as a friend or follower of Thomas of Woodstock, he remains a dominant figure throughout the play. One suspects that *Woodstock* was first written for a private performance at the real Sir Thomas Cheney's house on the Isle of Sheppey, Sir Thomas, of course, to play himself as his non-historical namesake in the play. This surmise is borne out by the fact that Woodstock's castle (according to Grafton--Holinshed calls it a house) at Plashey is taken as Plashey House in the play, and furthermore, Plashey, some twenty-five miles north of the Thames River in Essex, is moved to a spot "... near the Thames, circled round with trees", the latter a more apt description of Cheney's house on the Isle of Sheppey in the Thames River. Even "Arden woods", or the woods belonging to Sir Thomas Cheney's close friend, Master Arden of Faversham, and located south of the Thames surely, are moved to Essex on the north bank in a Trojan horse sort of masque. But here, if anywhere, geographical accuracy may be sacrificed to the poet's license with facts to please a patron.

The author wishes to thank Professor Robert Dilligan for programming assistance and for access to the facilities of the USC Computer Center. The permission to reproduce portions of the *Woodstock* manuscript, obtained from the British Museum, is gratefully acknowledged, as also is the permission from the Folger Shakespeare Library to reproduce handwriting specimens from the *Massacre at Paris* fragment.

27 Mustang Road  
Rancho Palos Verdes, California 90274  
U.S.A.

Louis ULE

## NOTES

1. E.K. Chambers, *The Elizabethan Stage*, Oxford, Clarendon Press, 1923, vol. IV, p. 42, where *Woodstock* is discussed under the name *I Richard the Second* because of its close relation to Shakespeare's *Richard II*. The *Woodstock* manuscript itself lacks a title page.
2. E.M. Thompson, *Shakespeare's Handwriting*, 1916. Thompson's authority, that Hand D in *Sir Thomas More* is Shakespeare's, is accepted by Chambers in his *William Shakespeare*, vol. I, pp. 498-515. S. Schoenbaum, in his *Internal Evidence and Elizabethan Dramatic Authorship*, Northwestern University Press, 1966, p. XIX, states : "On the basis of such [internal] evidence, Shakespeare's hand is now universally recognized to be present in *Sir Thomas More*".
3. L.A. Ule, "Cluster Analysis", ALLC Bulletin, 2, n° 3 (Michaelmas Term, 1974). This paper describes a method by which the N points of a frequency distribution curve are plotted as a single point in N-dimensional space, so that any number M of such frequency distribution curves of any quantifiable textual characteristic of M texts appear as M points in N-dimensional space. For values of N equal to or less than three, physical models can be constructed in which the texts so displayed tend to form clusters according to author, theme or genre. For larger values of N a purely numerical method, employing the Euclidean distance as a measure of closeness between these M points, is used to discover and list loose and tight clusters, each text or corresponding point taken in turn as the nucleus of a potential cluster. A graphical example is given in two dimensions of the relative frequency of use of three common English

words in twenty Elizabethan texts. A second paper, "The Use of CONSTAT in Authorship Investigations", presented at ICCH/2, USC, Los Angeles, April 1975, and to appear in a forthcoming issue of the ALLC Bulletin, contains a listing of the FORTRAN cluster analysis computer program and examples of typical computer output.

4. Christopher Marlowe's literary maturity at age twenty-three is evidenced by the following, oft quoted, passage from *1 Tamburlaine* : "If all the pens that euer poets held,/ Had fed the feeling of their maisters thoughts,/ And euey sweetnes that inspir'd their harts,/ Their minds, and muses on admyred theames :/ If all the heauenly Quintessence they still/ From their immortall flowers of Poesy,/ Wherein as in a myrroure we perceiue/ The highest reaches of a humaine wit./ If these had made one Poems period/ And all combin'd in Beauties worthinesse,/ Yet should ther houer in their restlesse heads,/ One thought, one grace, one woonder at the least,/ Which into words no vertue can digest". According to Swinburne this is "one of the noblest passages, perhaps indeed the noblest, in the literature of the world, ever written by one of the greatest masters of poetry in loving praise of the glorious delights and sublime submission to the everlasting limits of his art".
5. Wilhelmina P. Frijlinck, *The First Part of the Reign of King Richard the Second or Thomas of Woodstock*, The Malone Society Reprints, 1929, ed. W.W. Gregg.
6. A.P. Rossiter, *Woodstock, A Moral History*, 1946, Chatto and Windus, London.

7. F.S. Boas, *Shakespeare and the Universities*, D. Appleton and Company, New York, 1923, Chapters V and VII. The finding that this collection of plays was bequeathed to Dulwich College by William Cartwright the younger, appeared in an article by F.S. Boas in *The Library* in 1917.
  
8. This manuscript, a single sheet written on both sides, may not actually be in Marlowe's hand, though A.D. Wraight and Virginia Stern in their *In Search of Christopher Marlowe*, The Vanguard Press, Inc., New York, 1965, have attempted to show that it is consistent with the single extant example of Marlowe's signature.
  
9. For example, a series of astrological treatises and yearly almanacs by John Harvey of Cambridge made their appearance, beginning in 1583 with "An Astrological Discourse vpon the great and notable Coniunction of the two superiour Planets SATVRNE & IUPITER, which shall happen the 28. day of April 1583". R.B. McKerrow in his *The Works of Thomas Nashe*, vol. V, p. 167, notes : "Harvey's discourse seems to have awakened immense interest, and, among the vulgar at least, a good deal of perturbation. It must be remembered that it was less than three years from the earthquake of 1580, which had stirred popular imagination to a degree out of all proportion to its violence".
  
10. John Bakeless in his *The Tragical History of Christopher Marlowe*, Harvard University Press, 1942, devotes a whole chapter, "Marlowe and Shakespeare", to this problem, which he best summarizes as follows : "The abundance of Shakespeare's quotations, echoes, and allusions [from Marlowe] is especially important because he lets his other

literary contemporaries severely alone". (Vol. 2, p. 213).

11. Specimens of handwriting from the *Massacre at Paris* fragment are reproduced by permission of the Folger Shakespeare Library, Washington, D.C. and those from the *Woodstock* manuscript by permission of the British Museum. The specimens, some barely legible, were enlarged, traced in outline, and filled in black to improve their legibility, then reduced to their original size.
12. W.W. Gregg et al., *English Literary Autographs*, Oxford, 1925, 1928; S.A. Tannenbaum, *The Handwriting of the Renaissance*, The Columbia University Press, 1930; and G.E. Dawson and Laetitia Kennedy-Skipton, *Elizabethan Handwriting*, W.W. Norton & Company, New York, 1966. In addition to these references, the author consulted over 100 photostat copies of Elizabethan hands in his possession.
13. R. Dilligan and L. Ule, "The Mathematics of Style", in a letter to the Editor, *The Times Literary Supplement*, 22 October 1971, give some results on letter frequency distribution in Elizabethan texts.
14. G. Udny Yule, *The Statistical Study of Literary Vocabulary*, Cambridge University Press, 1944.
15. William Urry, personal correspondence. In a letter dated 23 May 1975, Professor Urry expressed his views on the authorship of *Arden* of Faversham as follows : "I find it impossible to escape Marlowe's authorship. If he didn't do it, who then did at that date ? 'Lord' Cheney



and his men were in and out of Canterbury, with close if second-hand contacts with the Marlowes. If Marlowe's name had been on the title page then no one would have questioned his authorship. I have left the matter open in a note in my book".

And and and And And and  
 B B B B  
 d. d d. d  
 Enter Enter  
 ended Landlord  
 possession possession  
 welcome welcome all your  
 yett yett

Figure 1. Handwriting specimens from the *Woodstock* manuscript, the *Massacre at Paris* fragment and Shakespeare's will.