

## **Cornelia, the pirated copy** by Hartmut Ilsemann

Ever since Thomas Kyd published his closet drama<sup>1</sup> *Cornelia* early in 1594, which he had translated from the French original *Cornélie* published in 1574 by Robert Garnier, there has never been any doubt about his authorship. Quite to the contrary, his dedication to the Countess of Sussex in which he bids for aristocratic patronage is very explicit:

I shall beseech your Honour to repaire, with the regarde of those so bitter times, and priuie broken passions that I endured in the writing it.’<sup>2</sup>

Kyd was to die a couple of months later, a consequence of brutal tortures instigated by the Privy Council in the attempt to find the source of mutinous libels that had been posted around London in May 1593. The writings that the authorities had found in Kyd’s possession belonged to Christopher Marlowe with whom Kyd shared lodgings, while they were both in their patron’s service. Marlowe, however, was killed in Deptford on 30 May, stabbed to death by fellow government agent Ingram Frizer.

Kyd was best known for his most popular play *The Spanish Tragedy* (1587), and it is certainly surprising that he now turned his attention to a play that was never meant to be performed. Was he just imitating the fashionable aristocratic tastes of Mary Sidney, Countess of Pembroke, whose own translation of Garnier’s *Marc Antoine* had been printed in 1592, as Curtis Perry remarked in his abstract of ‘The Uneasy Republicanism of Thomas Kyd’s *Cornelia*’?<sup>3</sup> Or was he actually involved in furthering the political thought of his time, exploring the limits of royal authority and the implementation of native liberties? The latter points would certainly have been part of his own experience that could be taken up again in the play and woven into the historical situation of a Roman Republic which was on the brink of mutating into the imperial rule of Caesar. A third option is now offered by new stylometric features in the *R Stylo* program known as Rolling Delta und Rolling Classify.<sup>4</sup>

Kyd was in dire straits after he had been released from torture and prison. Not only had he lost his patron, his estate was debt-ridden and his health was poor. A manuscript at hand for publication would have been more than welcome. In his dedication he qualifies the play as ‘small endeauours’, which critics understood as an acknowledgement of the weakness of the

---

<sup>1</sup> A play to be read rather than acted.

<sup>2</sup> *The works of Thomas Kyd, 1558-1594*, edited by Frederick S. Boas, (Oxford: Clarendon Press, 1901), p.102. Digitized by Google from the collections of Harvard University (<https://archive.org/details/worksthomaskyd00kydgoog>)

<sup>3</sup> From: Curtis Perry, “The Uneasy Republicanism of Thomas Kyd’s *Cornelia*,” *Criticism*, Volume 48, Number 4, Fall 2006, pp. 535-555 | 10.1353/crt.2008.0009 <https://muse.jhu.edu/article/231447#f3>, accessed on 28.01.2018

<sup>4</sup> Eder, M., Rybicki, J. and Kestemont, M. (2016). Stylometry with R: a package for computational text analysis. *R Journal* 8(1): 107-121. <<https://journal.r-project.org/archive/2016/RJ-2016-007/index.html>>

translation. But it may just as well have been a hint at material that had come into his possession and was not his own work.

To find possible relationships with other texts of the time, the carefully edited 1825 edition of *Cornelia*<sup>5</sup> was linked with a large number of reference texts. In these preliminary stages, plays by Chapman, Chettle, Greene, Munday, Lyly, Lodge, Peele, Rowley and Shakespeare could be eliminated and the following reference plays were left in the `primary_set` folder:

```
- kyd_soliman.txt (17867 words)
- kyd_spanpure.txt (17151 words)
- mar_tamburlain1.txt (17609 words)
- mar_tamburlain2.txt (17694 words)
- nashe_summer.txt (16740 words)
```

The file in the `exploration` folder was:

```
- kyd_cornelia.txt (15622 words)
```

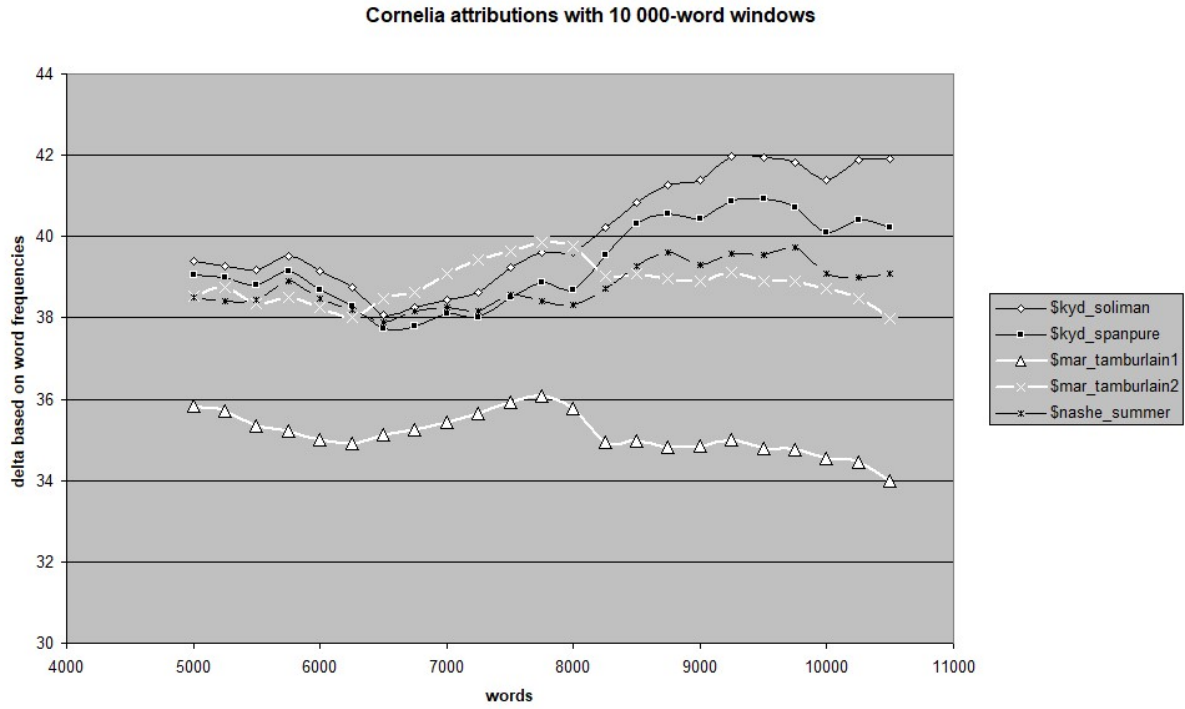
After Burrows had introduced Delta as a measure of the smallest stylistic difference in 2002, only sole-authored whole plays could be analysed, whereas Rolling Delta allowed for overlapping window sizes and was thus able to indicate collaborations.<sup>6</sup> In the case of *Cornelia* a 10.000-word window represents delta, so to speak, and attributes the play in its entirety to one author. The following three charts are based on such windows and variables as word frequencies and the frequencies of character bi- and trigrams (in brief: MF1W, MF2C and MF3C).<sup>7</sup>

---

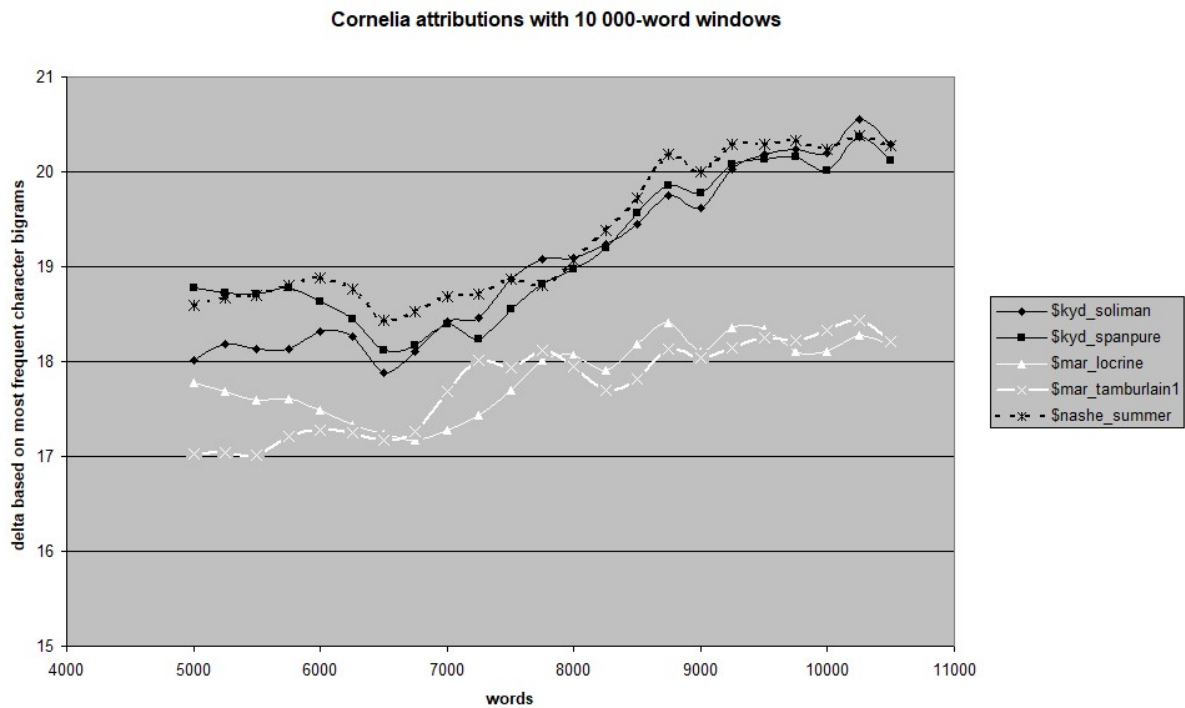
<sup>5</sup> Robert Dodsley, (1825), *A Select Collection of Old Plays (3rd ed.)*, vol. II, edited by John Payne Collier. London: Septimus Prowett, (This edition was chosen over Boas's 1901-edition for its modernised spelling to match the reference texts).

<sup>6</sup> Hoover quotes Burrows's original definition of Delta as 'the mean of the absolute differences between the z-scores for a set of word-variables in a given text-group and the z-scores for the same set of word-variables in a target text' in "Testing Burrows's Delta", *Literary and Linguistic Computing*, vol. 19, no 4, 2004, 453-475. Burrows had introduced Delta in: J. F. Burrows (2002). 'Delta': a measure of stylistic difference and a guide to likely authorship. *Literary and Linguistic Computing*, 17(3): 267-87

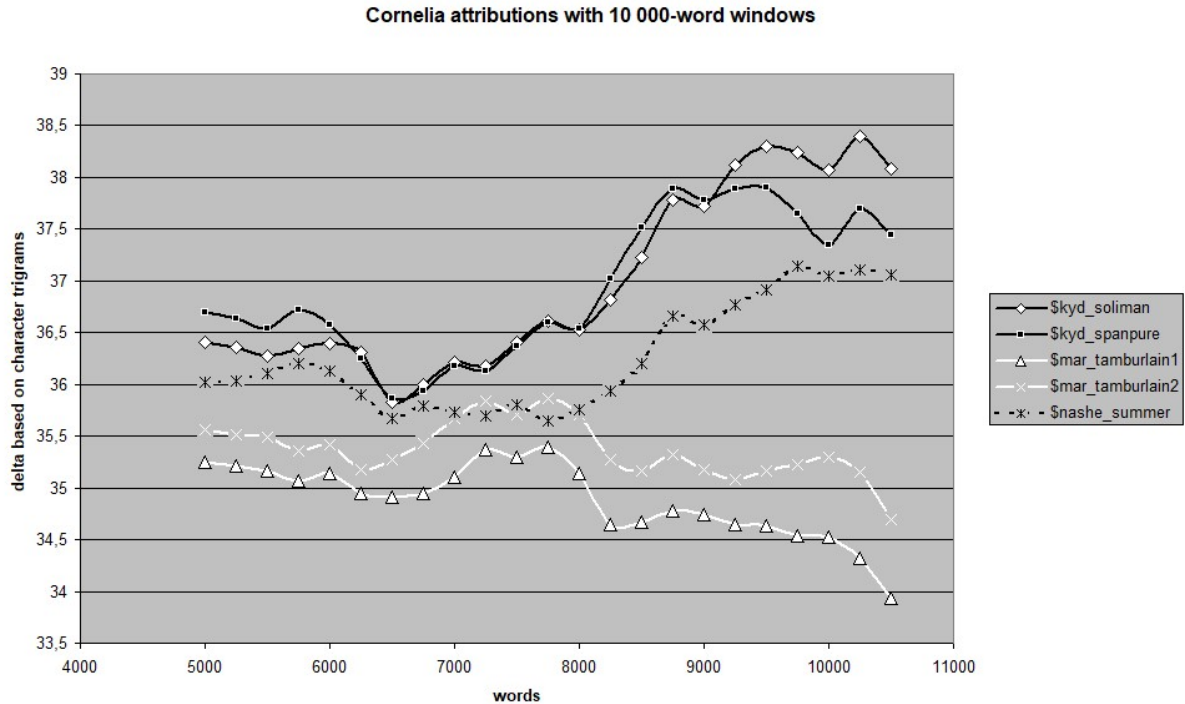
<sup>7</sup> Jack Grieve, (2007), "Quantitative Authorship Attribution: An Evaluation of Techniques," *Literary and Linguistic Computing*, vol. 22, no 3, 251-270



**Figure 1** *Cornelia* – delta attribution with MF1W



**Figure 2** *Cornelia* – delta attribution with MF2C



**Figure 3** *Cornelia* – delta attribution with MF3C

Consistently the three types of variables return *Tamburlaine 1* by Christopher Marlowe as the reference text with the lowest delta values.<sup>8</sup> The conclusion is obvious. Among the papers that Marlowe left behind must have been *Cornelia*, which would make Marlowe also the translator of the Garnier play. There is little doubt that he knew enough French, as he was operating as a secret agent in France during his repeated and prolonged absences from Cambridge.

The question that arises is rather whether Marlowe’s manuscript was left untouched by Kyd or if *Cornelia* was the result of collaborative endeavours.

Instead of many test trials and charts like those previously shown, an economical new presentation makes note of the best candidate only at each measuring point, i.e. the candidate with the lowest delta value, and the candidate in question is coded by a single capital letter, as, for example, K for Kyd, M for Marlowe, and N for Nashe.

The outcome can be seen in Table 1, a matrix that displays text segments of 250 words vertically (column A), beginning at 500 words, where B5 is the first measuring point of the 1000-word window. The first 1000 words of the play were attributed to Marlowe. The next window covers 1500 words and its first measuring point (C6) at 750 words returns Marlowe as well. 750 words is also the second measuring point of the 1000-word window marking words 251-1250 of the text. Accordingly, the first measuring point of the 5000-word window at 2500 words can be found at J13. As the window sizes grow horizontally, as indicated in

<sup>8</sup> The apocryphal play *The Tragedy of Lochrine* is normally not counted among the Marlowe corpus, but extensive investigations laid down in “Christopher Marlowe: Hype and Hoax” submitted to, and accepted by *Digital Scholarship in the Humanities* (2017) prove clearly a Marlowe provenance.



In terms of quantity, the matrix displays 1094 Marlowe cells, 253 by Kyd and 99 by Nashe. This alone would qualify Marlowe as the originator of the closet play, just as *delta* (here represented by the 10 000-word window at the end of each of the three series of tests differing by 500 words) had indicated before. But one would also have to take into account the differences between the three types of variable and between the windows of different size.<sup>9</sup> The default window size of *R Stylo* is 5000 words and Eder warns explicitly against smaller windows. He shifted his attention to sizes between 2500 and 5000 words when he examined 522 drama, prose and poetry texts from English, German, Polish, Hungarian and Latin literature.<sup>10</sup> In *Cornelia* we find almost exclusively Marlowe when Rolling Delta is based on the most frequent words in the texts. With the most frequent character bi- and trigrams, the window sizes return contradictory information, and it is Grieve's examination of variables (see footnote 7) which draws our attention to the attribution columns of MF3C. The number of variables is much higher and for this reason also more reliable. Here Nashe comes into play as a potential collaborator of Marlowe in II.1 (see column AF, where acts and scenes are marked, next to the final column AG with the cumulative word counts). On the other hand, there are continuous Kyd contributions covering all three sections around the end of III.1, until they are swallowed by the larger windows, in this case 4500, 5000, and 10000 with MF3C. A provisional conjecture might claim Marlowe and Nashe as co-authors of the play, as indicated in the columns and confirmed by empirical evidence, while allowing that Kyd contributed some revisions in light of the fact that Kyd's dedication claims his responsibility for 'small endeavours'.

It may well be that the actual scope of 'small endeavours' might even dwindle further if the so far anonymous play *The Tragedy of Lochrine* comes in as a reference text. Previous investigations established this play as a play by Marlowe (see Appendix). Many attributions of 250-word segments that could be claimed mostly by Marlowe's *Tamburlaine 1* now indicate *Lochrine* as the text with the lowest delta values. Table 2 accounts for these changes.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	
1	Window sizes and attributions: <i>Cornelia</i>																																	
2	T	1	1	2	2	3	3	4	4	5	10	1	2	2	3	3	4	4	5	10	1	2	2	3	3	4	4	5	10					
3	H	0	5	0	5	0	5	0	5	0	0	0	5	0	5	0	5	0	5	0	0	0	5	0	5	0	5	0	5	0	0			
4	T	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0				
5	O	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0					
6	500	M										M										M												
7	750	M	M									M	L									M	L											
8	1000	M	L	M								M	L	M								M	L	M										
9	1250	L	M	M	M							L	M	M	M							L	L	L	L									
10	1500	L	M	M	M	M						M	M	M	L	M						L	L	L	L	L								
11	1750	L	L	L	M	M	M					L	L	L	M	M						L	L	L	L	L	L					I. 1	1696	
12	2000	L	L	L	L	M	M	M				K	K	L	M	M	M					L	L	L	L	L	L	L	L					
13	2250	L	L	L	L	M	M	M	M			K	K	L	L	L	M	M	M			L	L	L	L	L	L	L	N					
14	2500	L	L	L	L	M	M	M	M	M		K	K	L	L	M	M	M	M			L	L	L	L	L	L	N	N					

<sup>9</sup> Maciej Eder, (2015), "Does size matter? Authorship attribution, small samples, big problem", *Digital Scholarship in the Humanities*, 30(2), 167-182

<sup>10</sup> Eder, 169, 181.

15	2750	L	L	K	M	M	M	M	M	M	K	K	L	M	L	M	M	M	M	N	L	L	L	L	L	N	N	N				
16	3000	L	L	K	K	M	M	M	M	M	K	L	M	L	M	M	M	M	M	K	L	L	L	L	L	N	L	L				
17	3250	K	L	K	K	M	M	M	M	M	N	M	L	K	M	M	M	M	M	K	L	L	K	L	N	L	L	L				
18	3500	K	N	N	M	M	M	M	M	M	M	N	M	L	M	M	M	M	M	K	N	N	N	N	L	L	L	L				
19	3750	K	N	N	M	M	M	M	M	M	K	K	M	M	M	M	M	M	M	K	K	N	N	N	L	L	L	L				
20	4000	N	M	K	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	N	N	N	N	N	L	L	L	L				
21	4250	N	M	M	K	M	M	M	M	M	M	M	M	M	M	M	M	M	M	N	N	N	N	N	L	L	L	L				
22	4500	N	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	N	N	N	L	L	L	L	L	L				
23	4750	L	K	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	N	N	L	L	L	L	N	L	L	II.1	4778		
24	5000	L	M	M	M	M	M	M	M	M	M	L	M	M	M	M	M	M	M	L	L	L	L	L	L	L	L	L				
25	5250	L	L	M	M	M	M	M	M	M	M	L	L	M	M	M	M	M	M	M	L	L	L	L	L	L	L	L	L			
26	5500	L	L	M	M	M	M	M	M	M	M	L	L	M	M	M	M	M	M	M	L	L	L	L	L	L	L	L	L			
27	5750	L	L	M	M	M	M	M	M	M	M	L	M	L	M	M	M	M	M	M	L	L	L	L	L	L	L	L	L			
28	6000	M	L	M	M	M	M	M	M	M	M	L	L	L	L	M	M	M	M	M	L	L	L	L	L	L	L	L	N	L		
29	6250	L	M	M	M	M	M	M	M	M	M	L	L	L	M	M	M	M	M	L	L	L	L	L	L	L	L	L	L			
30	6500	L	M	M	M	M	M	M	M	M	M	L	M	L	L	L	M	M	M	L	M	L	L	L	L	L	L	L	L			
31	6750	M	L	M	M	M	M	M	M	M	M	K	K	L	L	L	L	M	L	M	L	L	L	L	L	L	L	L	L			
32	7000	L	L	L	M	M	M	M	M	M	M	K	K	K	K	L	L	L	L	L	L	L	L	L	L	L	L	L	L			
33	7250	L	L	L	M	M	M	M	M	M	M	K	K	K	K	K	L	L	L	L	L	L	L	L	L	L	L	L	L			
34	7500	L	L	L	M	M	M	M	M	M	M	K	K	K	K	K	L	L	L	L	L	L	L	L	L	L	L	L	L			
35	7750	K	K	L	K	M	M	M	M	M	M	K	K	K	K	K	L	L	L	L	L	L	L	L	L	L	L	L	L			
36	8000	N	K	K	K	M	M	M	M	M	M	K	K	K	K	K	M	M	K	K	K	N	N	L	L	L	L	L	L	III.1	8001	
37	8250	K	K	K	K	M	M	M	M	M	M	K	K	K	K	K	K	M	L	K	K	N	N	L	L	L	L	L	L			
38	8500	K	K	K	K	M	M	M	M	M	M	K	K	K	K	K	L	M	K	N	N	N	N	L	L	L	L	L	L			
39	8750	K	K	K	K	M	M	M	M	M	M	N	K	K	K	K	L	M	N	N	N	N	L	L	L	L	L	L	L			
40	9000	M	K	M	M	M	M	M	M	M	M	N	K	K	M	K	K	K	K	K	N	N	L	L	L	L	L	L	L			
41	9250	K	M	M	M	M	M	M	M	M	M	K	M	L	M	M	M	K	K	M	N	M	L	L	L	L	L	L	L			
42	9500	M	M	M	M	M	M	M	M	M	M	K	M	L	M	L	K	L	M	N	L	L	L	L	L	L	L	L	L			
43	9750	N	M	M	M	M	M	M	M	M	M	M	L	M	L	L	L	L	K	L	M	L	L	L	L	L	L	L	L	IV.1	9764	
44	10000	L	L	M	M	M	M	M	M	M	M	L	M	L	M	L	L	L	L	L	L	L	L	L	L	L	L	L	L			
45	10250	L	M	M	M	M	M	M	M	M	M	L	L	M	L	L	L	L	L	L	L	L	L	L	L	L	L	L	N	L		
46	10500	L	L	M	M	M	M	M	M	M	M	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	N	N	L	
47	10750	L	K	L	M	M	M	M	M	M	M	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	N	L	L	
48	11000	K	L	N	L	M	M	M	M	M	M	K	L	L	L	L	L	L	L	K	L	L	L	L	L	L	L	M	M			
49	11250	K	K	K	N	M	M	M	M	M	M	K	L	L	L	L	M	L	L	N	N	N	L	L	L	L	M	L	M			
50	11500	N	K	N	M	N	M	M	M	M	M	K	K	K	L	M	L	M	L	N	N	N	N	N	M	M	M	M	M			
51	11750	K	N	N	M	M	M	M	M	M	M	K	K	K	M	M	M	M	M	N	N	N	N	L	M	M	M	M	M			
52	12000	M	N	N	M	M	M	M	M	M	M	L	K	K	M	M	L	M	M	N	N	N	N	M	M	M	M	M	M			
53	12250	N	M	M	M	M	M	M	M	M	M	K	M	M	M	M	M	M	M	N	N	M	M	M	M	M	M	M	M			
54	12500	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	L	L	N	M	M	M	M	M	M	M	M	L				
55	12750	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	L	M	M	M	M	M	M	M	M	M	M	M			
56	13000	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M			
57	13250	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M			
58	13500	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M			
59	13750	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M			
60	14000	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M			
61	14250	M	M	M	M	M	M	M	M	M	M	M	L	M	M					M	M	M	M									
62	14500	L	L	M								K	M	M						L	L	M										
63	14750	L	M									L	M							L	M											
64	15000	L										L								L												
65																																
66																																
67																																
68																																
69																																

L = anon. *The Tragedy of Locrine*  
 K = Kyd. *Soliman and Perseda*; *The Spanish Tragedy* (pure)  
 M = Marlowe. *Tamburlaine 1*; *Tamburlaine 2*  
 N = Nashe. *Summers Last Will and Testament*

Table 2 *Cornelia*

The number of cells in Table 2, i.e. text segments of 250 words, is of course identical, but the Marlowe reference texts now occupy 691 cells as compared to 1094 in Table 1, Kyd's number goes down to 155 instead of 253, Nashe loses 5 from 99 to 94 and the reference text *Locrine* now takes 506 segments from 1446 altogether. This means that due to the lower delta values of *Locrine* the Marlowe attributions have gained another 103 cells and dominate 83 % of the matrix, whereas in Table 1 Marlowe had almost 76 % of all cells. Together with the Nashe sections that may have gone into the composition of the play we encounter some 90 % of the text that was definitely not written by Kyd whose changes cover only approx. 10 %.

It is certainly an asset when independent methodologies come to the same or similar conclusions. Machine-learning algorithms are normally well suited to build up classifiers that can identify the author of the text in question. Rolling classify makes use of classifiers like *delta* (classic Burrowsian), *svm* (support vector machine) and *nsc* (nearest shrunken centroid),

all of which are part of the R Stylo package.<sup>11</sup> These supervised machine-learning classifications are combined with the idea of a sequential analysis in which segments of 5000 words are consecutively attributed with an overlap of 4750 words so that every 250-word section is given an attribution. The delta classification carried out on the basis of the most frequent character trigrams (MF3C) returns Marlowe exclusively. So do the exceptionally precise *svm* classifications. But there are other attributions as the list of classifications acknowledges. Once again the classifications are based on the frequencies of words (MF1W), character bigrams (MF2C) and character trigrams (MF3C), and each time the classifiers *delta*, *svm* and *nsc* are used.

> delta.mf1w\$classification.results

```
[1] nashe kyd kyd kyd kyd kyd kyd kyd kyd mar
[11] mar kyd kyd kyd kyd kyd kyd kyd kyd kyd
[21] mar mar mar mar mar mar mar mar mar mar
[31] mar mar mar mar mar mar mar mar mar mar
[41] mar
(total number of elements: 41)
```

> svm.mf1w\$classification.results

```
[1] mar mar mar mar mar mar mar mar mar mar mar mar mar mar mar
[15] mar mar mar mar mar mar mar mar mar mar mar mar mar mar mar
[29] mar mar mar mar mar mar mar mar mar mar mar mar mar mar mar
(total number of elements: 41)
```

> nsc.mf1w\$classification.results

```
[1] mar mar mar mar mar mar mar mar mar mar mar mar mar mar mar
[15] mar mar mar mar mar mar mar mar mar mar mar mar mar mar mar
[29] mar mar mar mar mar mar mar mar mar mar mar mar mar mar mar
(total number of elements: 41)
```

> delta.mf2c\$classification.results

```
[1] mar mar mar mar mar mar mar mar mar mar mar mar mar mar mar
[15] mar mar mar mar mar mar mar mar mar mar mar mar mar mar mar
[29] mar mar mar mar mar mar mar mar mar mar mar mar mar mar mar
(total number of elements: 41)
```

> svm.mf2c\$classification.results

```
[1] mar mar mar mar mar mar mar mar mar mar mar mar mar mar mar
[15] mar mar mar mar mar mar mar mar mar mar mar mar mar mar mar
[29] mar mar mar mar mar mar mar mar mar mar mar mar mar mar mar
(total number of elements: 41)
```

> nsc.mf2c\$classification.results

```
[1] mar mar mar mar mar mar mar mar mar mar
[11] mar mar mar mar mar mar mar mar mar mar
[21] mar mar mar mar nashe nashe nashe nashe kyd nashe
[31] nashe nashe mar mar mar mar mar mar mar mar
[41] mar
(total number of elements: 41)
```

> delta.mf3c\$classification.results

```
[1] mar mar mar mar mar mar mar mar mar mar mar mar mar mar mar
```

---

<sup>11</sup> A comprehensive description can be found in Eder, M. (2016b). "Rolling stylometry." *Digital Scholarship in the Humanities*, 31(3): 457-469. [pre-print] and in Eder's "Testing Rolling stylometry" <https://sites.google.com/site/computationalstylistics/projects/testing-rolling-stylometry>.

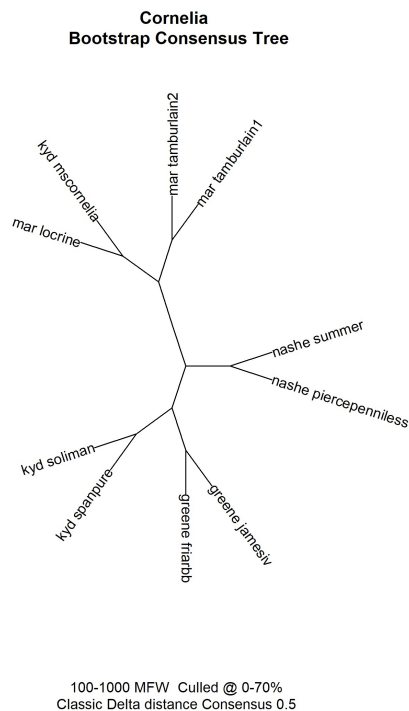




42	11000	M	M	M	M	M	M	M	M	M		
43	11250	M	M	M	M	M	M	M	M	M		
44	11500	M	M	M	M	M	M	M	M	M	IV.1	11384
45	11750	M	M	M	M	M	M	M	M	M		
46	12000	M	M	M	M	M	M	M	M	M		
47	12250	M	M	M	M	M	M	M	M	M		
48	12500	M	M	M	M	M	M	M	M	M		
49												
50												
51												
52												
53											V.1	14957

**Table 3** Classificatory attributions

The strongest impression comes from *svm* and *delta* attributions, which record Marlowe exclusively with MF2C and MF3C. With MF1W *delta* reports Kyd at the beginning of the play and *nsc* attributions have a number of Nashe signals. But some considerations have to be watched carefully. Eder for example reckons that *nsc* is ‘less hesitant in classification’<sup>12</sup>, that is, the decision level is much lower than in *delta* and *svm*. But it is also the mathematical



**Figure 4** Bootstrap Consensus Tree positioning of *Cornelia*

Other stylometric tools of the *R Stylo*<sup>13</sup> suite also confirm the Marlovian nature of *Cornelia*. *Stylo*'s bootstrap consensus tree places *Cornelia* together with other plays by Marlowe like the two *Tamburlaines* and the so far apocryphal *Tragedy of Lochrine* that Rolling Delta identified as a Marlowe play (see Figure 5, Appendix and footnote 8).

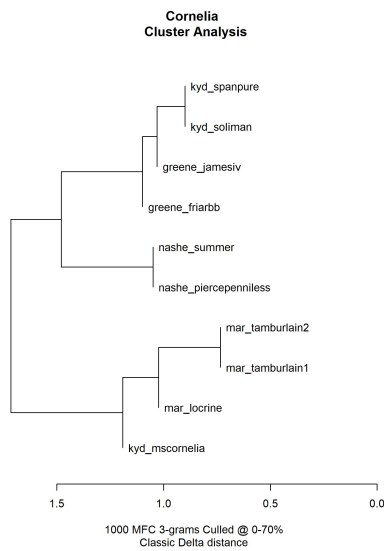
<sup>12</sup> Eder, M. (2016b), 460.

<sup>13</sup> Eder, M., Kestemont, M. and Rybicki, J. (2016). “Stylometry with R: A package for computational text analysis“. *R Journal*, 16(1): 107-121, <https://journal.r-project.org/archive/2016/RJ-2016-007/index.html>

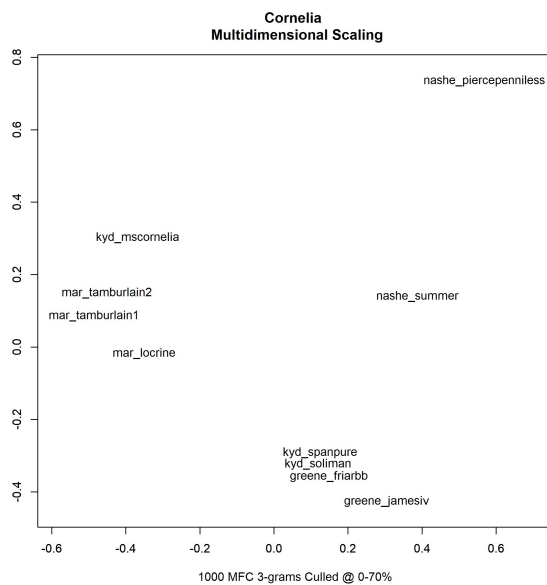
kernel of the classifiers which exerts a considerable influence on the results. Normally *delta* and *svm* are the reliable classifiers, and once again it is decisive when one takes into account the advantages that character bi- and trigrams have over word-based analyses. With these reservations, one can certainly conclude that the classifications too suggest that *Cornelia* is a Marlovian play which may have been marginally revised by Kyd.

Kyd's other plays are diametrically opposed in this arrangement that relied on word frequencies. Needless to say that character bi- and trigrams reproduce exactly the same relations.

Cluster analysis, too, produces the same array of plays with *Cornelia* located with Marlowe's plays rather than Kyd's. The variable employed here was MF3C (see Figure 6).



**Figure 5 Cluster Analysis of file set**



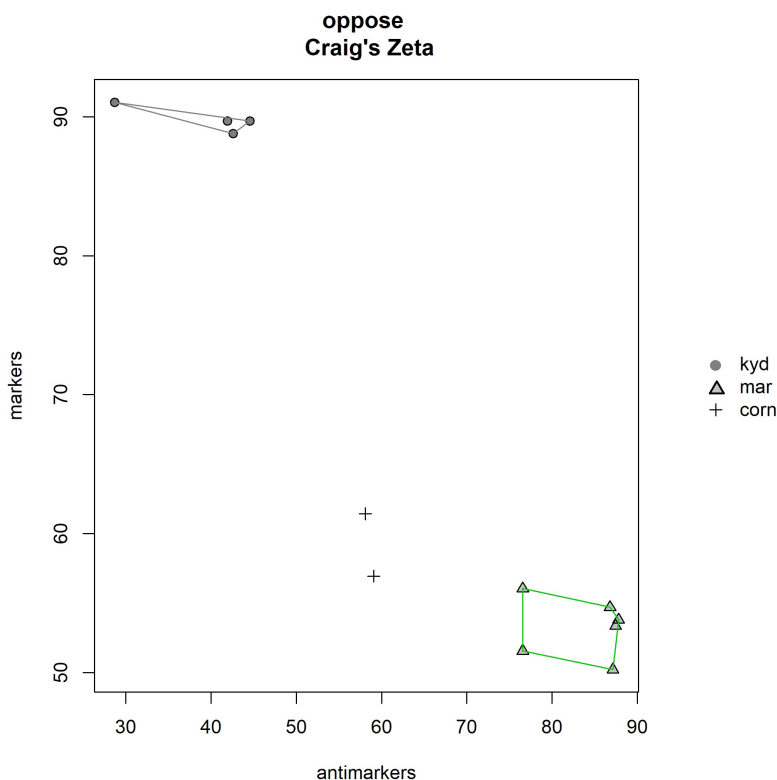
**Figure 6 Multidimensional Scaling**

Likewise, multidimensional scaling in Figure 7 places *Cornelia* closer to the *Tamburlaines*. An interesting result can be observed with the *oppose* function of *R Stylo* and the choice of Craig's Zeta, a formula that is based on Burrows's original Zeta<sup>14</sup>. Hoover gives its main feature as follows:

<sup>14</sup> Burrows, J. F. (2007). All the way through: testing for authorship in different frequency strata. *Literary and Linguistic Computing*, 22(1): 27-48.

“Zeta analysis excludes the extremely common words that have traditionally been the focus and concentrates on the middle of the word frequency spectrum. [...] Craig has developed an alternative version of Zeta that simultaneously creates sets of marker words and anti-marker words and has applied it impressively to Shakespeare authorship problems [...]. Although Craig's Zeta focuses on the same part of the word frequency spectrum as Zeta, its calculation and results are quite different. [...] It begins with two sets of texts divided into about equal-sized sections, then compares how many sections for each author contain each word, ignoring the frequencies of the words and concentrating on their consistency of appearance. The most natural comparison is between two authors, but it can be used to study any contrast.”<sup>15</sup>

Here the confrontation of Kyd and Marlowe produces the strange situation that *Cornelia* is placed closer to the domain of Marlowe’s words.



**Figure 7** *Cornelia* placement between the Marlowe and Kyd compounds

One crucial question remains at the end of this paper. How do the results fit with the endeavours of other authorship studies?<sup>16</sup> The answer is obvious: Not very well. It is equally clear that discrepancies do not arise from the careless handling of texts and materials. It is the chosen method that decides on the outcome, and here Gary Taylor, John V. Nance and

<sup>15</sup> Hoover, David L. (2010), “The Craig Zetra Spreadsheet,” *Digital Humanities*, <http://dh2010.cch.kcl.ac.uk/academic-programme/abstracts/papers/html/ab-659.html>, accessed on 19.03.18.

<sup>16</sup> A comprehensive and exemplary survey of the methodology of authorship attributions was provided by Gabriel Egan in ‘A History of Shakespearean Authorship Attribution’, *The New Oxford Authorship Companion*, ed. by Gary Taylor & Gabriel Egan, Oxford: Univ. Press, 2017, pp. 27-47

Keegan Cooper have only just recently ventured into the snares and pitfalls of attributions.<sup>17</sup> Their main target is Brian Vickers and his ‘advocacy of Kyd,’ who according to Vickers had written most of *Edward III*, *1 Henry VI*, ‘and all of *Arden of Faversham*, *Fair Em* and *King Leir*’ (p.146). In this respect Rolling Delta results are in accordance with their criticism. To break away from Vickers’s line of argument and to avoid self-defeating circularity, Taylor, Nance and Cooper established a couple of algorithms. Their first point, not to assume to know the author of the text in question, finds the full support of Rolling Delta procedures, in that a vast number of reference texts have been consulted, and it was in this way that Kyd’s *Cornelia* appeared, contrary to all expectations, to be of Marlovian provenance. To use databases and search engines that are publicly available certainly makes sense, particularly when results have to be checked. There is however one constraint with Rolling Delta, in that texts have undergone some changes. Speaker names, secondary text information and punctuation have been removed to exclude editorial influences. What is left are speeches alone in lower case letters. To identify the author of samples correctly is indeed one of the problems in authorship attributions, and here the danger of circularity is always present. Once a text has been wrongly attributed, its samples cannot be used to identify other texts. This is not only a problem of the Marlowe corpus, where most of his plays have been wrongly attributed, if one is to believe Rolling Delta results. Taylor, Nance and Cooper use a control sample from *Cornelia* to successfully identify Kyd’s writing, in this case in *Edward III*, which had been wrongly attributed to Kyd by Vickers. As long as their use of ‘two to four consecutive words (‘n-grams’) or juxtapositions of two or more semantically significant words within ten words of one another (‘collocations’)’ (p. 147) is confirmed by *Edward III* this would lead to a faulty attribution. Luckily the n-gram figures did not match, particularly not in the Mariner’s speech on which Vickers had relied. Likewise Shakespeare, Greene or Peele were unlikely to have written the speech. So in the end it was the unique verbal parallels and philological expertise that pointed to Marlowe. But the discrepancy remains in the present state of stylometric assessments, on the one hand we have ‘laboriously collected, detailed data’ focusing on a small amount of text and therefore not permitting inferences to be drawn about a full-length play (see also Egan and footnote 16), and on the other hand there is a new methodology that uses longer sole-authored reference texts permitting the minutest stylistic differences to show up, sometimes even in smaller sections of the target text.

---

<sup>17</sup> Taylor, G., Nance, J., & Cooper, K. (2017). Shakespeare and Who? Aeschylus, Edward III and Thomas Kyd. In P. Holland (Ed.), *Shakespeare Survey 70: Creating Shakespeare* (Shakespeare Survey, pp. 146-153). Cambridge: Cambridge University Press. doi:10.1017/9781108277648.015











