Hartmut Ilsemann

The Authorship of *George a Greene, the Pinner of Wakefield*.

Abstract

Finding the authorship of *George a Greene, the Pinner of Wakefield* can be seen within the wider context of detecting a collaborative or single-authored canon of early Shakespeare plays which, perhaps, he did not want to see as part of his dramatic work if we are to trust Heminges' and Condell's references to 'surrepetitious copies, maimed, and deformed by the frauds and stealthes of injurious imposters, that expos'd them' (First Folio preface). The fact that the remaining plays of the First Folio [are] 'absolute in their numbers as he [Shakespeare] conceived them,' has in some way a ring of self-amputation. The conclusion of the present study, based on the acknowledged methodological advances of non-traditional stylometric tools, contained in the *R Stylo* suite, is that *George a Greene, the Pinner of Wakefield* was originally written by William Shakespeare, even though the play may have been shortened for performances outside London with a reduced company of players during the plague years of 1592 and 1593.

Introduction

In 1825 Isaac Reed, Octavius Gilchrist and John Payne Collier published the newly edited and corrected volume 3 of Dodsley's *A Select Collection of Old Plays*,[1] of which the whole collection had been first published in twelve volumes in 1744. The volume starts off with the comedy *George a Green the Pinner of Wakefield*, and Collier reports in his introduction that the author of the play was unknown (p. 3). Its title page was given as:

A Pleasant conceyted Comedie of George a Greene the Pinner of Wakefield. As it was sundry times acted by the Servants of the Right Honorable, the Earl of Sussex. Imprinted at London, by Simon Stafford, for Cuthbert Burgy; and are to be sold at his Shop neare the Royall Exchange, 1599, 4to.(p. 48)

Further information comes from H. Dugdale Sykes, letting us know that the play 'was entered in the Stationers' Register by Cuthbert Burby on April 1, 1595, but that it was written not later than 1593 is shown by Henslowe's Diary, which records five performances by the Earl of Sussex's men between December 29 of that year and January 22 following" (p. 129). He then refers to a copy's title page of the 1599 quarto, which has two additional inscriptions in seventeenth-century handwriting, running thus:

(1) Written by . . . a minister who ac[ted] the piners pt in it himself. Teste W. Shakespea[re][2]

(2) Ed. Iuby saith that the play was made by Ro. Gree[ne]

Sykes is at the outset not fully convinced that Greene is actually the author of the play, since the first writer left a blank for the author's name, and Greene was not known to have

been a minister or a clerk. But nevertheless he comes to the conclusion that Greene must be the author.

> A careful scrutiny of the text has satisfied me that Greene was indeed the author of George a Greene, and my reasons for this conclusion I shall now give as shortly as possible (Sykes, p. 130).

Chambers too acquaints us with the notion that 'Greene's authorship has been very commonly accepted,' (p. 15) and furthermore mentions a number of ascriptions by Fleay and Oliphant, ranging from Greene and Peele to Lodge and a revision by Heywood. Chambers closes with the note:

> R. B. McKerrow thinks (M.S.C. p. 289) that the ' by Ro. Greene ' of the note may mean ' about Ro. Greene ' as a leading incident is apparently based on an episode of Greene's life. An allusion in I. i. 42 to *Tamburlaine* gives an anterior limit of date" (p. 15).

Accordingly Pervez Rizvi in his data bank of common n-grams and collocations dates *George a Greene* at 1587, but Early Print's metadata give the source date as 1591. It is self-evident that all these qualifications provide an incentive to tackle the authorship problem with the *R Stylo* suite (Eder, Kestemont, Rybicki), particularly with *Rolling Delta*, *Rolling Classify* and the *General Imposters* method (GI).

Stylometry

The advantage of these stylometric tools is certainly that they do not rely on editorial evidence and subjective comparisons of contents, motifs, parallels and the like. Their main target is the choice of vocabulary and textual structures. The first task in these procedures is to identify the reference texts that have the smallest stylistic difference from the target text. This alone was not only time consuming, but also a despairing undertaking as one potential author after the other had to be excluded. John Day whose play *The Blind Beggar of Bednal Green* had the lowest delta values was much too young when the play was performed. So he must be seen as a borrower. Other reference texts by Peele, Chapman and Munday were not single-authored and would have to be erased if the programme had determined them. In the end reference texts were used that contained many anonymous, and perhaps Shakespeare plays of the time which could be left disregarded in filtering out known and suitable authors of plays:

1. anon_arden1592.txt (19975 words)
2. anon_ashrew1594.txt (12273 words)
3. anon_edwardiii1594.txt (19316 words)
4. anon_fairem1590.txt (11606 words)
5. anon_famvicth5.txt (12616 words)
6. anon_ironside1590.txt (15257 words)
7. anon_kingleir1594.txt (22953 words)
8. anon_more1592.txt (19871 words)
9. anon_mucedorus1590.txt (12377 words)
10. anon_oldcastle1600.txt (21366 words)
11. anon_troublereign1591.txt (24243 words)
12. anon_truetragrichiii1594.txt (19794 words)
13. anon_woodstock1591.txt (27764 words)
14. chap_bussydambois1607.txt (21380 words)
15. chap_bussyrevenge1612.txt (19993 words)
16. chettle_hoffman.txt (19822 words)
17. greene_alphonsus.txt (14937 words)
18. greene_friarbb.txt (16746 words)
19. greene_jamesiv1591.txt (20190 words)
20. greene_orlando.txt (11130 words)
21. greene_Selimus.txt (18698 words)
22. kyd_soliman.txt (17890 words)
23. kyd_spantrag.txt (22590 words)
24. lodge_lookingglass.txt (19605 words)

25. lodge_mariusscilla1590.txt (19332 words)
26. lyly_campaspe1583.txt (12499 words)
27. lyly_motherBombie1588.txt (16675 words)
28. mar_tamburlain1.txt (17609 words)
29. mar_tamburlain2.txt (17694 words)
30. mun_deathh1601.txt (22902 words)
31. mun_downfall1598.txt (20469 words)
32. mun_kentcumberms.txt (13592 words)
33. nashe_piercepenniless.txt (31134 words)
34. nashe_summer.txt (16740 words)
35. peele_alcazar1591.txt (11082 words)
36. peele_arraignment.txt (10335 words)
37. peele_david1599.txt (15108 words)
38. peele_edward1.txt (21864 words)
39. peele_mixpoems.txt (10413 words)
40. peele_oldwives1595.txt (7707 words)
41. row_whenysee.txt (24723 words)
42. sack_gorboduc.txt (14448 words)
43. shak_2henry4.txt (26675 words)
44. shak_antcleo1606.txt (24469 words)
45. shak_asyou1599.txt (21734 words)
46. shak_caesar1599.txt (19508 words)
47. shak_coriolan1608.txt (27338 words)
48. shak_cymbeline1610.txt (27699 words)
49. shak_errors1594.txt (14693 words)
50. shak_hamlet.txt (30345 words)
51. shak_hamlet1600.txt (30364 words)
52. shak_henryv1598.txt (26292 words)
53. shak_henryviii1613.txt (23664 words)
54. shak_john1596.txt (20904 words)
55. shak_lear1606.txt (26127 words)
56. shak_lovelab1594.txt (21545 words)
57. shak_macbeth.txt (17113 words)
58. shak_merchant1596.txt (21027 words)
59. shak_mfm1603.txt (21786 words)
60. shak_midsum1595.txt (16464 words)
61. shak_pericles1607.txt (18303 words)
62. shak_rapelucrece.txt (15031 words)
63. shak_richii1595.txt (22392 words)
64. shak_richiii1592.txt (29125 words)
65. shak_romjul.txt (24704 words)
66. shak_romjul1595.txt (24704 words)
67. shak_shrew1590.txt (20911 words)
68. shak_tempest1611.txt (16557 words)
69. shak_thnight.txt (19504 words)
70. shak_timon1605.txt (18390 words)
71. shak_titus1592.txt (20359 words)
72. shak_troilus1602.txt (26241 words)
73. shak_twokins1613.txt (23653 words)
74. shak_verona1590.txt (17272 words)
75. shak_windsor1597.txt (23404 words)
76. shak_winters1609.txt (25575 words)
77. sidney_marcantonie.txt (14484 words)
78. wilkins_misenfmarriage.txt (23844 words)
79. wilson_3ladieslondon1584.txt (19001 words)[3]

With *George a Greene* (9060 words)[4] in the secondary_set and the 79 reference texts above in the primary_set windows of 5000 words were tested with character trigrams (mf3c)[5] and with an overlap of 250 words so that a row of seventeen measurements for each reference text came into being. The matrix was copied into a spreadsheet and for each column of window centroids the three lowest values were highlighted. They were stylistically closest to *George a Greene*. There was however no indication of Greene, Munday or Peele. Almost exclusively Shakespeare plays were selected by the program as Fig. 1 shows. Both *Arden of Faverham*, a Shakespeare play according the New Oxford Shakespeare edition of 2016, and *Sir John Oldcastle*, proven in 2017 with *R Stylo* as Shakespeare's play (Ilsemann, 2017) were best suited in their stylistic closeness. When these two plays were disregarded and the next closest plays were shown (Fig. 2), there were only Shakespeare plays.

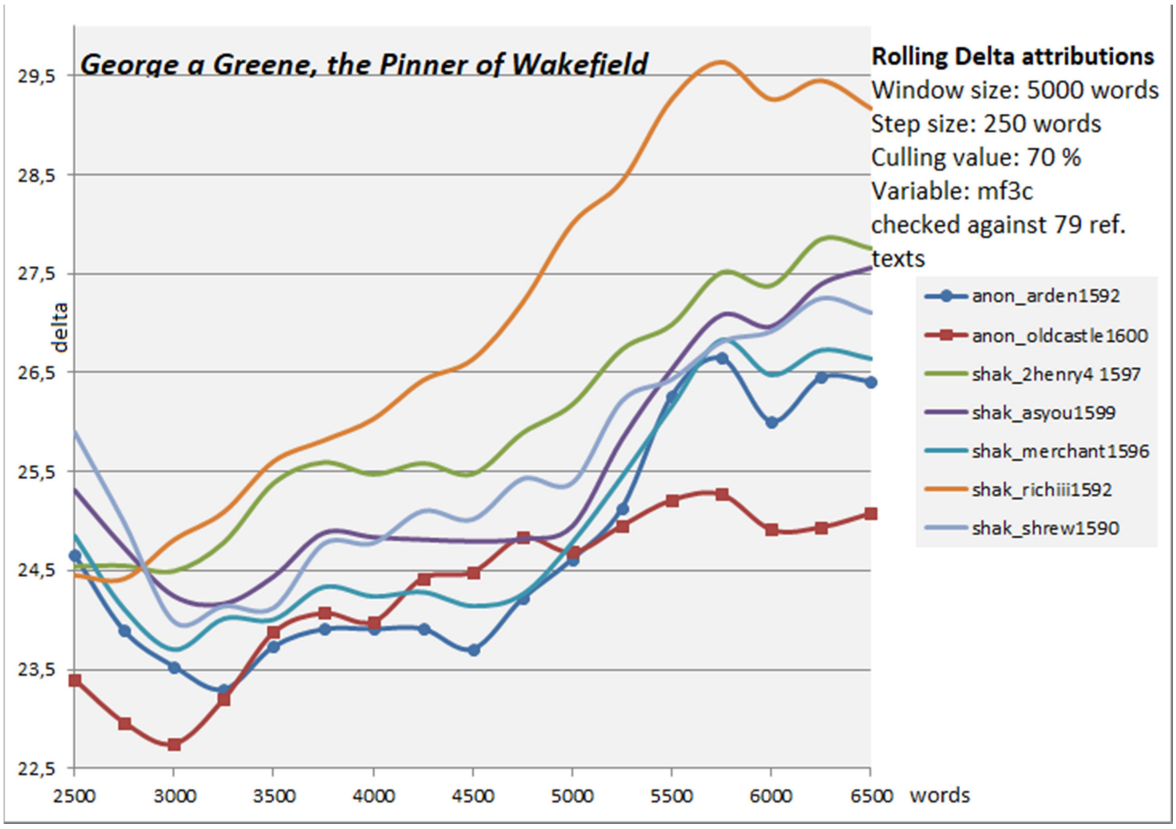

Fig. 1 ***Rolling Delta*** **results with contemporary reference texts**

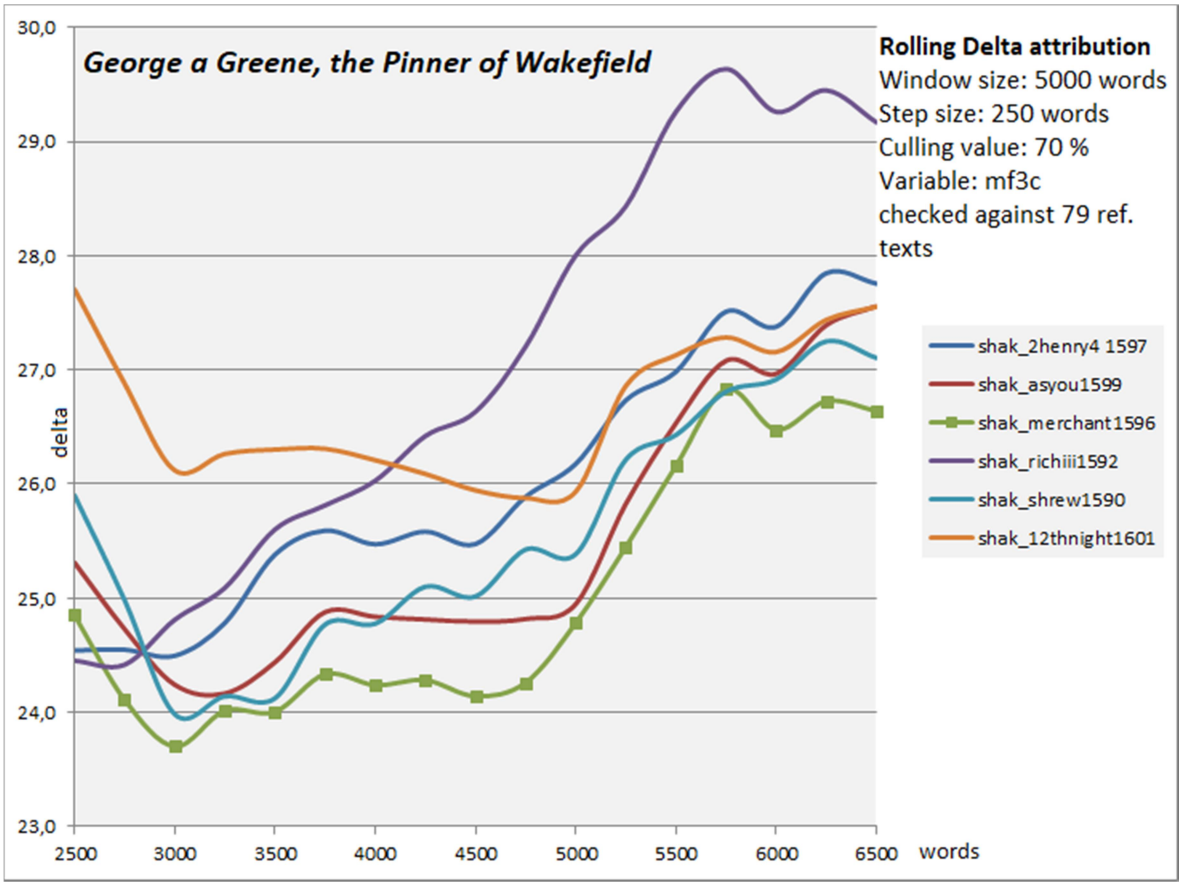

Fig. 2 ***Rolling Delta*** **evaluation of 79 reference texts**

Table 1 Figure 1 displayed as a table

	A	B	C	D	E	F	G	H	I	J
1	words	***George a Greene, the Pinner of Wake-***							**Scenes**	**Words**
2	0	***field***								
3	250									
4	500	**Rolling Delta attribution with mf3c**							I.1	418
5	750	Window size: 5000 words								
6	1000	Step size: 250 words							I.2	992
7	1250	Culling value: 70 %							I.3	1173
8	1500	checked against 79 reference								
9	1750	texts							I.4	1777
10	2000									
11	2250								II.1	2459
12	2500	24,7	23,4	24,5	25,3	24,9	24,5	25,9	II.2	2529
13	2750	23,9	23,0	24,5	24,7	24,1	24,4	25,0		
14	3000	23,5	22,7	24,5	24,2	23,7	24,8	24,0		
15	3250	23,3	23,2	24,8	24,2	24,0	25,1	24,1		
16	3500	23,7	23,9	25,4	24,4	24,0	25,6	24,1		
17	3750	23,9	24,1	25,6	24,9	24,3	25,8	24,8		
18	4000	23,9	24,0	25,5	24,8	24,2	26,0	24,8		
19	4250	23,9	24,4	25,6	24,8	24,3	26,4	25,1	II.3	4237
20	4500	23,7	24,5	25,5	24,8	24,1	26,6	25,0	III.1	4569
21	4750	24,2	24,8	25,9	24,8	24,3	27,2	25,4		
22	5000	24,6	24,7	26,2	24,9	24,8	28,0	25,4		
23	5250	25,1	25,0	26,7	25,8	25,5	28,4	26,2	III.2	5220
24	5500	26,3	25,2	27,0	26,5	26,2	29,3	26,4	III.3	5557
25	5750	26,6	25,3	27,5	27,1	26,8	29,6	26,8		
26	6000	26,0	24,9	27,4	27,0	26,5	29,3	26,9		
27	6250	26,4	24,9	27,8	27,4	26,7	29,4	27,3	IV.1	6388
28	6500	26,4	25,1	27,8	27,6	26,6	29,2	27,1		
29	6750	B	C	D	E	F	G	H	IV.2	6746
30	7000	7	10						IV.3	7019
31	7250	8	4			4	1			
32	7500	1	2	2	1	11		2	IV.4	7481
33	7750							%		
34	8000	B =	anon_arden1592				7			
35	8250	C =	anon_oldcastle1600				10	100		
36	8500	D =	shak_2henry4 1597							
37	8750	E =	shak_asyou1599							
38	9000	F =	shak_merchant1596						V.1	9060
39		G =	shak_richiii1592							
40		H =	shak_shrew1590							

Table 2 Figure 2 displayed as a table

	A	B	C	D	E	F	G	H	I
1	words	***George a Greene, the Pinner of***						**Scenes**	**Words**
2	0	***Wakefield***							
3	250								
		Rolling Delta attribution with							
4	500	**mf3c**						I.1	418
5	750		Window size: 5000 words						
6	1000		Step size: 250 words					I.2	992
7	1250		Culling value: 70 %					I.3	1173
			checked against 79						
8	1500		reference						
9	1750		texts					I.4	1777
10	2000								
11	2250							II.1	2459
12	2500	24,5	25,3	24,9	24,5	25,9	27,7	II.2	2529
13	2750	24,5	24,7	24,1	24,4	25,0	26,9		
14	3000	24,5	24,2	23,7	24,8	24,0	26,1		
15	3250	24,8	24,2	24,0	25,1	24,1	26,3		
16	3500	25,4	24,4	24,0	25,6	24,1	26,3		
17	3750	25,6	24,9	24,3	25,8	24,8	26,3		
18	4000	25,5	24,8	24,2	26,0	24,8	26,2		
19	4250	25,6	24,8	24,3	26,4	25,1	26,1	II.3	4237
20	4500	25,5	24,8	24,1	26,6	25,0	25,9	III.1	4569
21	4750	25,9	24,8	24,3	27,2	25,4	25,9		
22	5000	26,2	24,9	24,8	28,0	25,4	25,9		
23	5250	26,7	25,8	25,5	28,4	26,2	26,9	III.2	5220
24	5500	27,0	26,5	26,2	29,3	26,4	27,1	III.3	5557
25	5750	27,5	27,1	26,8	29,6	26,8	27,3		
26	6000	27,4	27,0	26,5	29,3	26,9	27,2		
27	6250	27,8	27,4	26,7	29,4	27,3	27,4	IV.1	6388
28	6500	27,8	27,6	26,6	29,2	27,1	27,6		
29	6750	B	C	D	E	F	G	IV.2	6746
30	7000			15	1	1		IV.3	7019
31	7250	1	5	1	1	9			
32	7500	1	9	1		5	1	IV.4	7481
33	7750								
34	8000	B =	shak_2henry4 1597						
35	8250	C =	shak_asyou1599						
36	8500	D =	shak_merchant1596				15		
37	8750	E =	shak_richiii1592				1		
			shak_shrew159						
38	9000	F =	0				1	V.1	9060
39		G =	shak_12thnight1601						
40							100	%	

In a further step the analysis was extended to window sizes between 1000 and 5000 words at a distance of 1000 words, and moreover the variables words (mf1w), character bigrams (mf2c) and character trigrams (mf3c) were used. For each window measurement the lowest delta attribution was transferred into the table below displaying window sizes and type of variables horizontally and the sequential structure of the play vertically. Each 250-word segment is given an attribution.

Table 3 *Rolling Delta* attribution with mf1w, mf2c and mf3c

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1		**Rolling Delta attributions in *George a***																
2		***Greene, the Pinner of Wakefield***																
3		**Window sizes:**															**Scenes**	**Words**
4		**1000**	**2000**	**3000**	**4000**	**5000**	**1000**	**2000**	**3000**	**4000**	**5000**	**1000**	**2000**	**3000**	**4000**	**5000**		
5	words																	
6	500	S	**mf1w**				S	**mf2c**				S	**mf3c**				I.1	418
7	750	C					S					S						
8	1000	C	K				G	S				S	G				I.2	992
9	1250	K	K				G	S				K	S				I.3	1173
10	1500	K	K	K			K	S	S			K	K	S				
11	1750	K	K	K			S	S	S			S	K	S			I.4	1777
12	2000	K	K	S	S		S	S	S	S		K	S	S	S			
13	2250	K	K	S	S		S	S	S	S		S	S	S	S		II.1	2459
14	2500	K	S	S	S	S	S	S	S	S	S	S	S	S	S	S	II.2	2529
15	2750	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S		
16	3000	S	S	S	S	S	S	S	S	S	S	P	S	S	S	S		
17	3250	S	S	S	S	S	P	S	S	S	S	P	S	S	S	S		
18	3500	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S		
19	3750	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S		
20	4000	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S		
21	4250	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	II.3	4237
22	4500	S	S	S	S	S	P	S	S	S	S	S	S	S	S	S	III.1	4569
23	4750	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S		
24	5000	S	S	S	S	S	S	S	S	S	S	M	S	S	S	S		
25	5250	S	S	S	S	S	S	S	S	S	S	M	S	S	S	S	III.2	5220
26	5500	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	III.3	5557
27	5750	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S		
28	6000	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S		
29	6250	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	IV.1	6388
30	6500	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S		
31	6750	S	S	S	S		S	S	S	S		S	S	S	S		IV.2	6746
32	7000	S	S	S	S		S	S	S	S		S	S	S	S		IV.3	7019
33	7250	S	S	S			S	S	S			S	S	S				
34	7500	S	S	S			M	S	S			S	S	S			IV.4	7481
35	7750	P	S				M	S				S	S					
36	8000	P	S				S	S				S	S					
37	8250	P					S					S						
38	8500	S					S					S						

39	8750		
40	9000	V.1	9060

All in all 375 text segments were given assignments of which about 90 % (Table 4, B3) were related to Shakespeare, 266 of them referred to *Arden of Faversham*, a domestic play that was officially registered as a play by Shakespeare in the New Oxford Shakespeare in 2016. The 4000 and 5000-word windows yielded a 100% Shakespeare attribution following Eder's prediction that smaller windows are unreliable. And yet, with mf3c the 3000-word window already returns Shakespeare (E3) and with mf2c it is already the 2000-word window (D3).

Table 4 Percentage of attributions with outliers in the small windows

	A	B	C	D	E
1	**percentages**		**mf1w**	**mf2c**	**mf3c**
2	windows	**all**	**4000**	**2000**	**3000**
3	**Shakespeare**	90,7	100	100	100
4	Chettle	0,8	0	0	0
5	Kyd	5,3	0	0	0
6	**Greene**	1,1	0	0	0
7	Munday	1,6	0	0	0
8	**Peele**	0,5	0	0	0

Rolling Classify provides further evidence of Shakespearean authorship (Table 5). In terms of methodology three different classifiers were employed, nearest shrunken centroid (nsc), support vector machine (svm) and the classical delta classifier. Furthermore the classifications

Table 5 *Rolling Classify* results with nsc, svm and the delta classifier

	A	B	C	D	E	F	G	H	I	J
1		mf1w			mf2c			mf3c		
2	Worte	**nsc**	**svm**	**delta**	**nsc**	**svm**	**delta**	**nsc**	**svm**	**delta**
3	2500	kyd	kyd	greene	shak	shak	shak	shak	shak	shak
4	2750	kyd	kyd	nashe	shak	shak	shak	shak	shak	shak
5	3000	kyd	kyd	kyd	shak	shak	shak	shak	shak	shak
6	3250	shak	kyd	kyd	shak	shak	shak	shak	shak	shak
7	3500	shak	shak	kyd	shak	shak	shak	shak	shak	shak
8	3750	shak	kyd	kyd	shak	shak	shak	shak	shak	shak
9	4000	shak	kyd	kyd	shak	shak	shak	shak	shak	shak
10	4250	shak	kyd	kyd	shak	shak	shak	shak	shak	shak
11	4500	shak	kyd	nashe	shak	shak	shak	shak	shak	shak
12	4750	shak	kyd	nashe	shak	shak	shak	shak	shak	shak
13	5000	shak	kyd	nashe	shak	shak	shak	shak	shak	shak
14	5250	shak	kyd	nashe	shak	shak	shak	shak	shak	shak

15	5500	shak	kyd	nashe	shak	shak	mun	shak	shak	shak
16	5750	shak	kyd	nashe	shak	shak	mun	shak	shak	shak
17	6000	shak	kyd	nashe	shak	shak	mun	shak	shak	shak
18	6250	shak	kyd	kyd	shak	shak	shak	shak	shak	shak

were carried out with words (mf1w in columns B,C,D), character bigrams (mf2c in columns E,F,G) and character trigrams (mf3c in columns H,I,J).

The chosen window size was 5000 words with an overlap of 4750 words so that the window centroids were given an attribution in every 250-word segment. The reference texts were the same as in the *Rolling Delta* procedure, and once again *Arden of Faversham* was employed as Shakespeare's play. There is clear evidence that character trigrams are more precise than words alone. Svm, too, should preferably be used, as Maciej Eder reports of a very high decision level in contrast to nsc which is very classification friendly (Eder, p. 460). Classifications are normally at their best when a choice between two authors is at stake. The higher the number of reference texts the larger becomes the number of outliers so that it makes always sense to use another test which was presented by Eder in his blog 'Authorship verification with the package 'stylo'' of the Computational Stylistics group in 2018. The *General Imposters* Method (GI) was introduced by Koppel and Winter (2014) and Kestemont applied it to the study of Julius Caesar's disputed writings (Kestemont et al., 2016). GI is seen as an additional check on similarities in writing styles and is claimed to be a second verification system. It goes beyond the simple assessment of similarity, but aims to state whether two documents are significantly more similar to each other than they are to other documents. The prerequisites necessary to use the function `imposters()` have been described by Eder as follows:

> It assumes that all the texts to be analysed are already pre-processed and represented in a form of a matrix with frequencies of features (usually words). The function contrasts, in several iterations, a text in question against (1) some texts written by possible candidates to authorship, or the authors that are suspected of being the actual author, and (2) a selection of 'imposters", or the authors that could not have written the text to be assessed. Consequently, a given candidate's class is assigned a score between 0 and 1 (Eder, 2018).

According to Eder, the reasonable assumption of the procedure is that any result above 0.5 can be seen as a successful verification of authorship. The classifier available is delta, but other classifiers are in the process of being prepared. Eder added however two more distance measures, Cosine delta (Wu), developed by the Würzburg computational stylistics group, and Ruzicka metrics (Ru). The latter consumes a very high computation time, but is regarded as highly reliable.

As Table 6 reveals, the tests were carried out with words and character trigrams. Lines 3 and 18 reproduce the attributions for the text in question, vertically followed by the other reference texts and horizontally we find the display of authors. It is no surprise that the same authors are highlighted which could also be found in previous investigations with *Rolling Delta* and *Rolling Classify*. Once again it seems to be the character trigram evaluation which is more precise and within the three measurements delta, Wu and Ru the latter which consumes a very high computation time contains clearer results, an observation resulting from a large number of tests that have been carried out lately with a variety of plays and their

Table 6 GI evaluation of George a Greene with the delta classifier, Würzburg distance (Wu) and Ruzicka metrics (Ru)

	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
1	**Frequencies of**	delta								Wu								Ru							
2	**Words**	**anon**	**chettle**	**greene**	**kyd**	**mar**	**mun**	**nashe**	**shak**	**anon**	**chettle**	**greene**	**kyd**	**mar**	**mun**	**nashe**	**shak**	**anon**	**chettle**	**greene**	**kyd**	**mar**	**mun**	**nashe**	**shak**
3	anon_georgegreene.txt		0.00	**0.72**	0.12	0.00	**0.75**	0.07	0.34		0.00	**0.54**	0.03	0.00	0.03	0.00	0.13		0.05	0.02	0.20	0.00	0.19	0.00	**0.90**
4	chettle_hoffman.txt	0.00	**1.00**	0.00	0.38	0.00	0.00	0.01	0.47	0.00	**1.00**	0.00	0.15	0.00	0.00	0.00	0.35	0.01	**1.00**	0.00	0.27	0.00	0.00	0.01	0.49
5	greene_friarbb.txt	**0.55**	0.00	**0.85**	0.28	0.00	0.00	0.05	0.07	0.48	0.00	**0.99**	0.09	0.03	0.00	0.00	0.01	0.03	0.10	**0.88**	0.21	0.00	0.00	0.12	0.45
6	greene_orlando.txt	0.25	0.02	**1.00**	0.37	0.15	0.00	0.00	0.00	0.17	0.00	**0.99**	0.11	0.36	0.00	0.00	0.00	0.00	0.20	**0.95**	0.47	0.13	0.00	0.00	0.04
7	kyd_soliman.txt	0.00	0.08	0.08	**1.00**	0.14	0.00	0.04	0.49	0.02	0.00	0.12	**0.76**	0.13	0.00	0.00	0.04	0.03	0.19	0.04	**0.94**	0.00	0.00	0.00	**0.55**
8	kyd_spantrag.txt	0.00	0.14	0.08	**0.92**	0.01	0.25	0.01	0.46	0.02	0.00	0.07	**0.91**	0.01	0.00	0.00	0.09	0.01	0.34	0.01	**1.00**	0.00	0.02	0.00	0.26
9	mar_tamburlain1.txt	0.01	0.01	0.11	0.42	**1.00**	0.00	0.09	0.00	0.01	0.00	0.48	0.22	**1.00**	0.00	0.00	0.00	0.00	0.05	0.43	0.23	**1.00**	0.00	0.01	0.00
10	mar_tamburlain2.txt	0.03	0.00	0.27	0.44	**1.00**	0.00	0.06	0.01	0.00	0.00	0.47	0.31	**1.00**	0.00	0.00	0.00	0.00	0.02	0.43	0.35	**1.00**	0.00	0.01	0.00
11	mun_kentcumberms.txt	0.43	0.00	0.01	0.35	0.00	**1.00**	0.01	0.14	**0.53**	0.00	0.02	0.12	0.00	**1.00**	0.00	0.14	0.34	0.00	0.00	0.20	0.00	**1.00**	0.00	0.36
12	nashe_summers.txt	0.04	0.01	0.01	0.05	0.00	0.03	**1.00**	**0.50**	0.08	0.00	0.02	0.02	0.00	0.00	**1.00**	0.44	0.00	0.03	0.07	0.08	0.00	0.00	**1.00**	0.45
13	shak_romjul1595.txt	0.02	0.39	0.05	0.36	0.00	0.08	0.15	**1.00**	0.03	0.00	0.02	0.04	0.00	0.00	0.00	**1.00**	0.00	0.37	0.02	0.45	0.00	0.00	0.00	**0.82**
14	shak_thnight.txt	0.09	0.08	0.00	0.42	0.00	0.14	0.20	**1.00**	0.13	0.00	0.00	0.05	0.00	0.00	0.02	**1.00**	0.30	0.09	0.00	0.19	0.00	0.00	0.11	**1.00**
15																									
16	**Frequencies of**	delta								Wu								Ru							
17	**character trigrams**	**anon**	**chettle**	**greene**	**kyd**	**mar**	**mun**	**nashe**	**shak**	**anon**	**chettle**	**greene**	**kyd**	**mar**	**mun**	**nashe**	**shak**	**anon**	**chettle**	**greene**	**kyd**	**mar**	**mun**	**nashe**	**shak**
18	anon_georgegreene.txt		0.03	0.00	0.13	0.00	**0.74**	0.02	**0.88**		0.00	0.02	0.01	0.00	0.24	0.00	0.46		0.03	0.00	0.07	0.00	0.18	0.01	**1.00**
19	chettle_hoffman.txt	0.00	**1.00**	0.00	0.37	0.00	0.00	0.00	0.41	0.00	**1.00**	0.00	0.23	0.00	0.00	0.00	0.36	0.00	**1.00**	0.00	0.33	0.00	0.00	0.01	**0.50**
20	greene_friarbb.txt	0.00	0.05	**0.80**	0.33	0.00	0.00	0.47	0.45	0.16	0.00	**1.00**	0.04	0.03	0.00	0.00	0.00	0.00	0.02	**0.75**	0.20	0.00	0.00	0.37	0.38
21	greene_orlando.txt	0.00	0.16	**0.95**	**0.57**	0.11	0.00	0.00	0.04	0.01	0.00	**0.75**	0.16	0.47	0.00	0.00	0.00	0.00	0.05	**0.99**	**0.53**	0.02	0.00	0.02	0.01
22	kyd_soliman.txt	0.00	0.36	0.00	**1.00**	0.00	0.00	0.07	0.35	0.01	0.00	0.05	**0.86**	0.11	0.00	0.00	0.05	0.00	0.11	0.01	**0.88**	0.00	0.00	0.12	**0.61**
23	kyd_spantrag.txt	0.00	**0.60**	0.00	**0.98**	0.00	0.00	0.03	0.17	0.00	0.00	0.05	**0.92**	0.01	0.00	0.00	0.06	0.00	0.37	0.01	**1.00**	0.00	0.00	0.01	0.14
24	mar_tamburlain1.txt	0.00	0.10	0.09	0.49	**1.00**	0.00	0.08	0.00	0.01	0.00	0.47	0.21	**1.00**	0.00	0.00	0.00	0.00	0.02	0.33	0.39	**1.00**	0.00	0.08	0.00
25	mar_tamburlain2.txt	0.00	0.03	0.04	0.44	**1.00**	0.00	0.18	0.01	0.00	0.00	0.46	0.18	**1.00**	0.00	0.01	0.00	0.00	0.00	0.09	0.24	**1.00**	0.00	0.28	0.00
26	mun_kentcumberms.txt	0.00	0.14	0.00	0.17	0.00	**1.00**	0.00	**0.60**	**0.57**	0.00	0.00	0.00	0.00	**1.00**	0.00	0.27	0.02	0.05	0.00	0.02	0.00	**1.00**	0.00	**0.58**
27	nashe_summers.txt	0.00	0.15	0.01	0.13	0.00	0.00	**1.00**	**0.50**	0.00	0.00	0.06	0.01	0.01	0.00	**1.00**	0.17	0.00	0.00	0.08	0.13	0.00	0.00	**1.00**	0.45
28	shak_romjul1595.txt	0.00	**0.56**	0.00	0.21	0.00	0.00	0.05	**0.98**	0.10	0.01	0.00	0.04	0.00	0.00	0.00	**1.00**	0.00	0.38	0.00	0.16	0.00	0.00	0.05	**0.98**
29	shak_thnight.txt	0.00	**0.56**	0.00	0.16	0.00	0.01	0.26	**1.00**	0.18	0.01	0.00	0.00	0.00	0.01	0.00	**1.00**	0.00	0.37	0.00	0.06	0.00	0.01	0.18	**1.00**

authors. Greene, Munday and Shakespeare are the names that were highlighted, but apart from the sober grades of Wu, which records only known authorships with mf3c (J18-Q29) the highest figures are those of Shakespeare whose columns seem to indicate a large amount of stylistic variety and interrelationship with other authors (I,Y). Even though the main message is obvious, there is doubt about the reliability of figures above 0.5. Does every value above 0.5 indicate authorship? This is questionable, and in fact the latest development is an optimized procedure which checks the grey area of doubtful attributions (Table 7). Jan Rybicki developed a so far unpublished script which gives the boundaries of the grey area. Values above the upper boundary (column C) indicate authorship, values below the lower boundary (column B) exclude authorship. With regard to *George a Greene* we find that Greene (column F) does nowhere qualify for authorship. Instead it is the Würzburg distance (Cosine delta) in lines 4 and 9 which gives priority to Munday with words (j4) and to Lyly

with character trigrams (H8). Classic delta, too, opts for Munday with mf1w (J3) and mf3c (J8).

Table 7 enlarged GI test of *George a Greene, the Pinner of Wakefield*

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	***George a Greene, the Pinner of Wakefield*, GI tests with delta, Würzburg distance and Ruzicka metrics**												
2	**mf1w**	low	high		**chettle**	**greene**	**kyd**		**mar**	**mun**	**nashe**	**shak**	
3	delta	0.21	0.78		0.01	0.38	0		0	1	0	0.47	
4	wu	0.1	0.73		0	0.3	0		0	0.91	0	0.42	
5	ru	0.25	0.64		0.06	0.04	0.12		0	0.2	0	1	
6													
7	**mf3c**	low	high	**chap**	**chettle**	**greene**	**kyd**	**lyly**	**mar**	**mun**	**nashe**	**shak**	
8	delta	0.15	0.61	0	0.01	0.28	0.01	0.29	0	1	0	0.09	
9	wu	0	0.81	0	0	0.2	0	0.81	0	0.58	0	0.02	
10	ru	0	0.78	0	0.08	0.04	0.12	0	0	0.27	0.02	0.99	

The calculation that correlates most with previous results is that of the Ruzicka metric (L5, L10). Rybicki's script also provided a diagram which is displayed below (Fig. 3.)

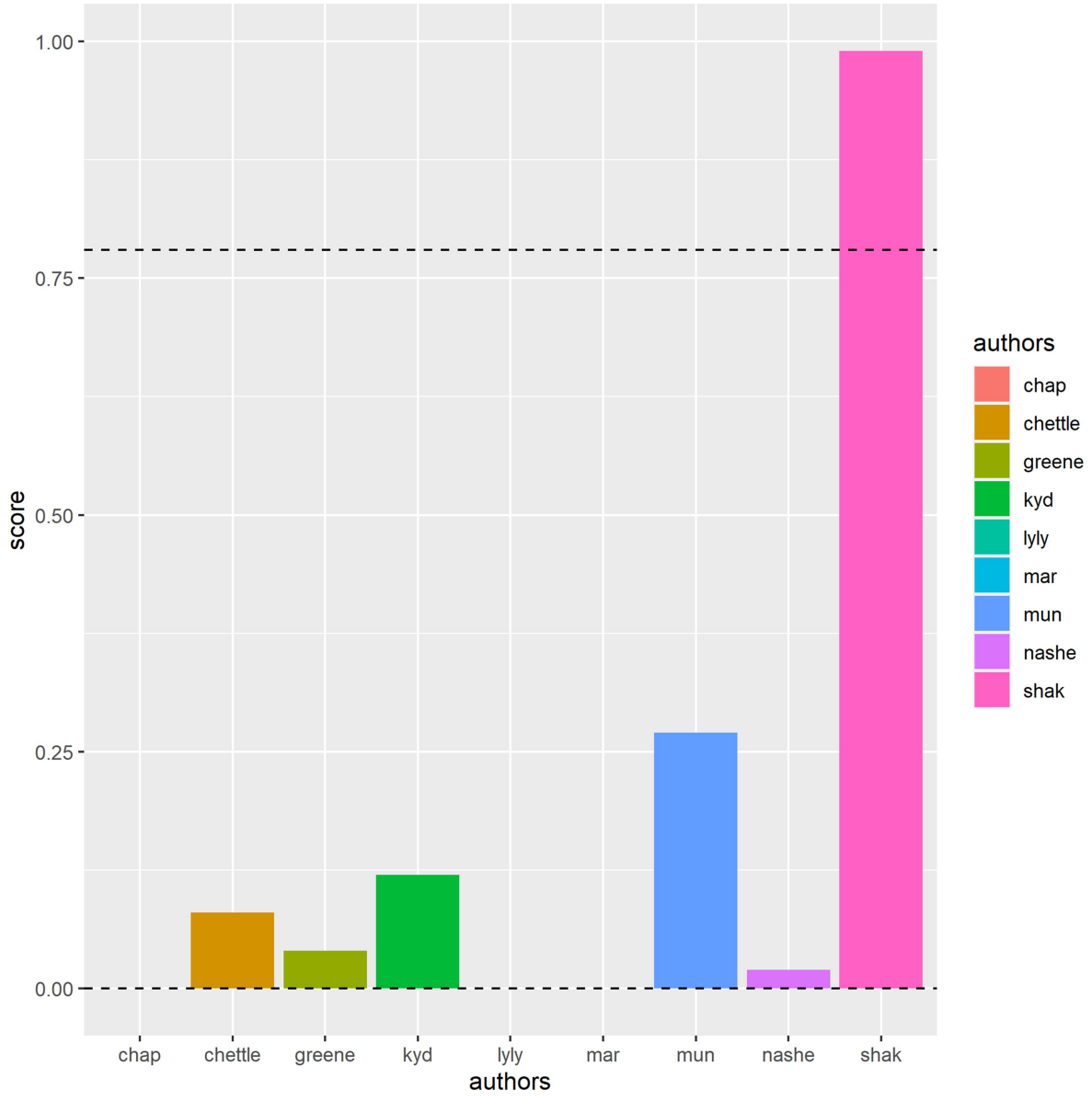

Fig. 3 optimized GI attribution of *George a Greene* with Ruzicka metrics and mf3c

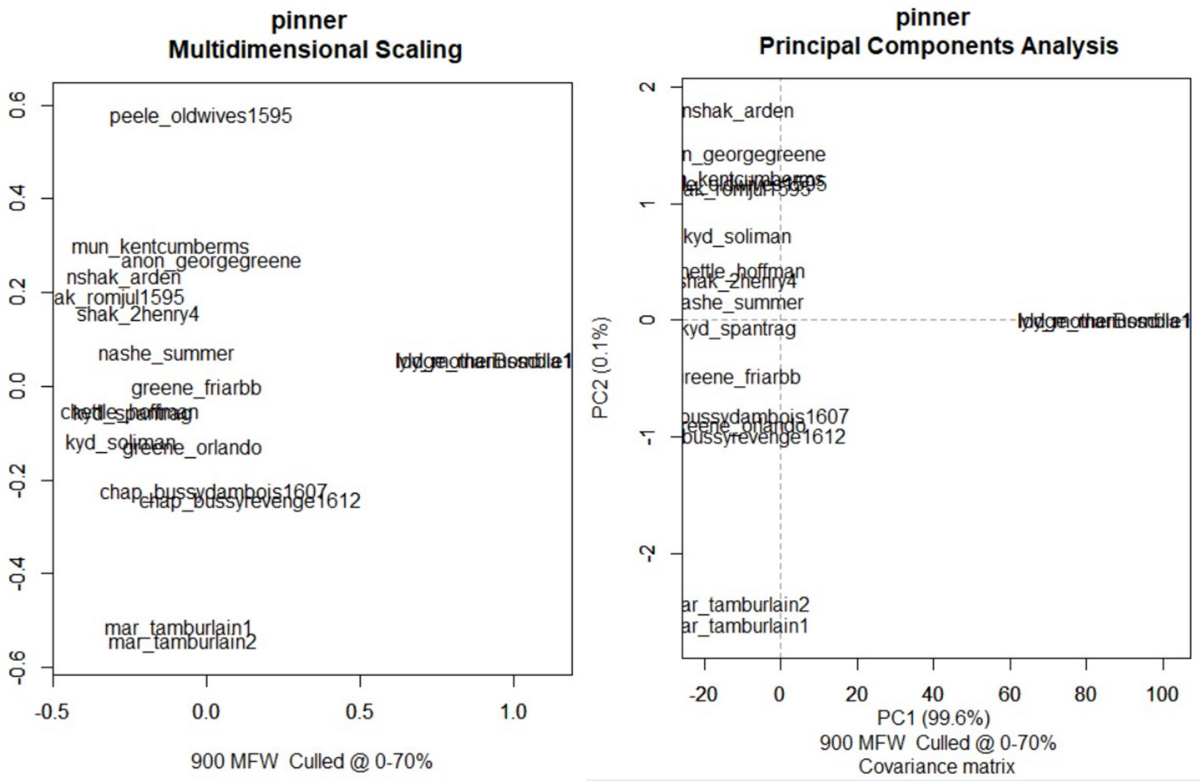

Fig. 4 MD and PCA ratings of vocabulary

One of the assets of *R Stylo* is its capacity of visualizing stylistic closeness and distances between reference texts. Both Multidimensional Scaling (MD) and Principal Component Analysis (PCA) provide evidence of *George a Greene*'s position between Shakespeare's *Arden of Faversham* and Munday's *John a Kent and John a Cumber*. MD also shows further Shakespearean environment with *Romeo and Juliet* and *2 Henry IV*. Cluster analysis (CA) too shows interrelationships between plays and by highlighting the course of edges it stresses the strength of attraction between plays (see Fig. 5). Apparently *George a Greene* has the strongest link with *Arden of Faversham* which again is similarly linked with *Romeo and Juliet*. The latter has strong connections with *2 Henry IV*. But one should not overlook that *George a Greene* is also associated with *Romeo and Juliet* and *The Old Wives' Tale* which relates again to *2 Henry IV* (see Fig. 5).

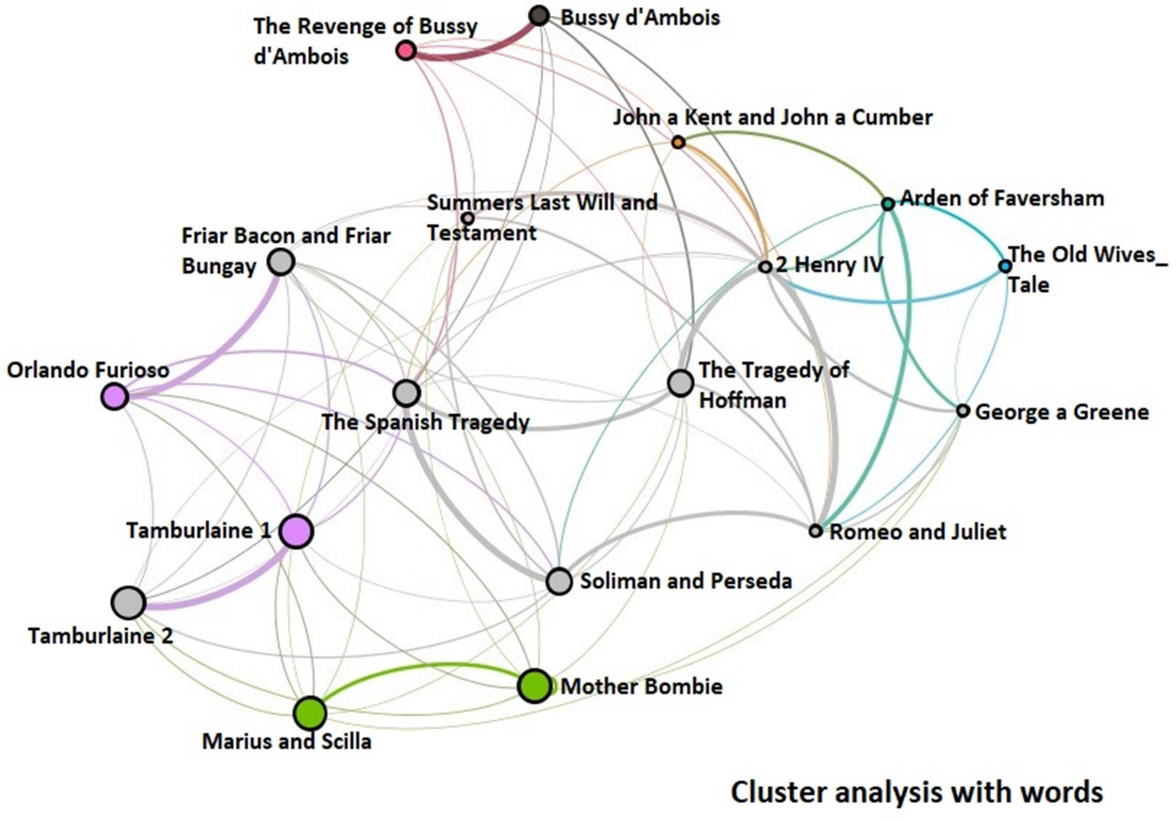

Fig. 5 CA links between *George a Greene* and neighbouring plays.

Evaluation

It is obvious that methodologies must also be discussed in the evaluation of results. When Sykes spoke out in favour of Greene he based his judgement on words like *blithe, bonny, mickle* and *hie* to name just a few of them. These he rated as typical of Greene's plays and poems and found them also in *George a Greene*. His counts were certainly correct, but at the time of his examination, in 1931, there were no computerised concordances. Otherwise he would have noticed that for example *hie* was used by Shakespeare in *Romeo and Juliet* alone eight times, in *Richard III* four times. In all examples the distribution was much wider than Sykes had assumed. Likewise N-grams and collocations had a high priority in the nineties of the last century and were often used in questions of authorship. They are definitely impressive to look at, but as Pervez Rizvi's data bank shows, they have too many interrelations with all sorts of plays and cannot be used as proof of authorship. When Burrows at the beginning of the millennium developed delta, function words were en vogue as stylistic discriminators. They could unquestionably hint at similarities between texts, but there were problems. Delta, in turn, was introduced by John Burrows in 2002 as a measure of stylistic difference and as a guide to likely authorship. For this, Burrows had developed a process that went beyond those previously used. Until then, among other things, the frequency of function words with their standard deviations had been used as a means of comparison between authors. Burrows, however, used so-called z-scores. This approach was based on the observation that the

frequency of occurrences in each word list decreases rapidly and the difference from the mean of the frequency increases with each word. To ensure that the rapidly decreasing frequency of words is equally included in the rating, the z-scores were calculated by dividing the difference between the mean and the actual frequency by the standard deviation. The respective result has a positive or negative value, depending on whether the word is above or below the mean. The absolute difference between the z-scores of the search text and the reference texts then gives the delta value, which is the expression of the stylistic difference between the texts. A prerequisite for working with Delta is of course that the appropriate reference texts are available. If they are missing, other texts and authors are displayed, but this brings forth incorrect results. In the following years, the usability of the delta method was confirmed, even if the functioning of the algorithm was not entirely clear. Improving suggestions came in 2004 from D. L. Hoover, who found that not the entire vocabulary of search text and reference texts should be used, but only about 70%, because the rest of the vocabulary represents rare and idiosyncratic words regarding very specific content. A culling value of 70%, on the other hand, harmonizes the results through improved comparability. The next step to improve quality was the investigation by Jack Grieve from 2007, which tested thirty-nine variables for their suitability to provide reliable information about authorship. In addition to the frequency of words, character bi- and especially character trigrams performed best. A short example may show the statistical advantage of the latter two variables. Let us take the most common word in many English texts, namely the direct article '_the_'. As a word, we list one variable that is determined by the preceding and the following blank. Noted as character bigrams we receive the sequences '_t', 'th', 'he', and 'e_', i.e. four variables, and with character trigrams we can use the last letter of the previous word and the first letter of the following word (here marked by x). The frequencies are now formed from 'x_t', '_th', 'the', 'he_' and 'e_x'. In practice, the effect is that with word variables there are around 90 words to be evaluated within a thousand words, whereas with character bigrams there are around 280 and with letter trigrams there are even around 750 variables. The statistical advantage is unique, especially since our everyday experience shows that about 1200 samples are needed for a coherent election forecast.

Qualitatively, the link across words gives access to an author's unconsciously designed diction, a skill that goes far beyond simple word frequencies. In contrast to authorship studies, which only worked with frequencies and in which the function words used provided a small number of strong discriminators, Burrows's Delta resorted to relatively weak discriminators, but in disproportionately large numbers. As a result, a figure was given out at the end of the procedure for each reference text, the lowest of which denoted the play with the smallest stylistic distance. And yet, Delta could only be meaningful if it was ensured that not only the research text but also the reference texts used were clearly assigned to an author. It was absolutely impossible to adequately grasp a collaborative text, and one had to use a trick by using old guesswork and comparing parts of the text separately with Delta. That this was associated with a susceptibility to errors is shown by the role of the reference text *Edward II*,

which was rated as Marlowe's play but actually originated from Kyd and Shakespeare. (see ‘Christopher Marlowe: Hype and Hoax,” Table 15).

The remedy was in place since about 2013 with the implementation of *Rolling Delta*. A word window of a certain size was pushed through the entire text with a selectable overlap and a delta measurement was taken each time. A measurement curve was created for each reference text. The text with the lowest measurement curve had the smallest stylistic difference from the search text. In contrast to this asset, particularly with mf3c, GI, MD and PCA with their references to texts by Greene, Lyly, Munday and Peele originate predominantly from static inventories. The one-time access to a certain set of words of whole texts prevents differentiation as it is possible by advancing word windows in a text. The conclusion from these methodological discussions and the results of the various tools employed in this study can only be that *George a Greene, the Pinner of Wakefield* was originally written by William Shakespeare, even though the play may have been shortened for performances outside London with a reduced company of players during the plague years of 1592 and 1593. That the Shakespeare corpus was with some certainty much larger can also be inferred from the fact that a playwright like Thomas Heywood claimed to have had ‘an entire hand or at least a maine finger in two hundred and twenty plays’ (Wikipaedia).

References

Burrows, J. F. (2002a). ‘Delta”: a measure of stylistic difference and a guide to likely authorship. Literary and Linguistic Computing, 17(3): 267–87.

Chambers, E. K. (1923). The Elizabethan Stage, vol. 4, Oxford: University Press, 1923, p. 15

Dodsley, Robert, ed. (1744). *A Select Collection of Old Plays* (1st ed.). London: Robert Dodsley. 12 Volumes

Reed, Isaac, Gilchrist, Octavius and Collier, John Payne, eds. (1825). Robert Dodsley. A Select Collection of Old Plays, London: Septimus Prowett

Eder, Maciej (2016). Rolling stylometry. Digital Scholarship in the Humanities, 31(3): 457-469.

Eder, Maciej (2013). Does size matter? Authorship attribution, short samples, big problem. Literary and Linguistic Computing, 28, Advance Access published 14.11.2013 (doi: 10.1093/llc/fqt066).

Eder, M., Kestemont, M. and Rybicki, J. (2013). Stylometry with R: a suite of tools. In Digital Humanities 2013: Conference Abstracts. University of Nebraska-Lincoln, NE, pp. 487–89.

Heminge, John and Condell, Henrie, Preface to The First Folio (1623), http://www.shakespeare-online.com/biography/firstfolio.html, accessed 08.03.2022.

Hoover, D. L. (2004). Testing Burrows’s delta. Literary and Linguistic Computing, 19(4): 453–75.

Ilsemann, Hartmut (2017), 'The Two Oldcastles of London', Digital Scholarship in the Humanities, 32(4): 788–796; doi: 10.1093/llc/fqw039

Ilsemann, Hartmut (2018), 'Christopher Marlowe: Hype and Hoax', *Digital Scholarship in the Humanities,* 33(4): 788–820, https://doi.org/10.1093/llc/fqy001

Kestemont, M., Stover, J., Koppel, M., Karsdorp, F. and Daelemans, W. (2016). Authenticating the writings of Julius Caesar. Expert Systems with Applications, 63: 86–96.

Koppel, M. and Winter, Y. (2014). Determining if two documents are written by the same author. Journal of the Association for Information Science and Technology, 65(1): 178–87 doi:10.1002/asi.22954. http://dx.doi.org/10.1002/asi.22954.

Pennel, Charles Alexander (1962). A critical edition of *George a Greene, the Pinner of Wakefield*. Dissertation, University of Illinois, University Microfilms, Inc. Ann Arbour, Michigan.

Sykes, H. Dugdale (1931) 'Robert Greene and George a Greene, the Pinner of Wakefield," The Review of English Studies, vol. 7, Nr. 26, April 1931, p. 129

Notes

[1] Robert Dodsley. *A Select Collection of Old Plays,* ed. by the late Isaac Reed, Octavius Gilchrist and John Payne Collier, London: Septimus Prowett, 1825

[2] Sir George Buc is said to have noted on the title page of *George a Greene, the Pinner of Wakefield* (1599), that he had consulted Shakespeare on its authorship. There is no specific source, but the line insinuates that William Shakespeare told Buc the play had been written by a playwright who was also a minister and who had acted the pinders part in the play.

[3] Text files begin with an (abbreviated) author name, followed by an underscore and a short play name. After the full stop the type of document is given, here: txt

[4] The play belongs to a number of short plays and Alexander Pennel in his *Critical Edition* of the play (1962) provided substantial evidence that the 1599 quarto was based on a report of an abridged performance of a non-extant, longer play (Pennel, p. 8). This may well refer to the years 1592/3 when the London theatres were closed due to the bubonic plague and theatre companies were forced to tour the country.

[5] Variables were denoted with abridgements, in which "mf" stands for "most frequent", "c" for characters, "w" for words, thus a six-word collocation would be "mf6w".