HARTMUT ILSEMANN

Addressing the Need to Revisit Authorship Attributions in English Renaissance Drama

This contribution is supposed to serve as an incentive to re-examine authorship attributions of English Renaissance plays. It starts from the assumption that cultural traditions, once they have come into existence, are long-lived and tenacious. It is only due to modern non-traditional stylometry with its new Rolling Delta and Rolling Classify features as well as the General Imposters Method that revisions of mistaken authorship ascriptions can be effected and become accepted.

In Shakespearean scholarship for centuries the first Folio edition of 1623 has been the gold standard, so to speak, for what was to be considered a Shakespeare text. Thirty-six plays constituted Shakespeare's oeuvre, to which *Pericles* and *The Two Noble Kinsmen* were later added. In comparison with the dramatist, actor and stage-manager William Shakespeare, the playwright Thomas Heywood claimed to have had "an entire hand or at least a maine finger in two hundred and twenty plays." (Clark, 219) The question unavoidably arose, as to whether Shakespeare had possibly written more plays than those known to posterity. The evolutionist Eric Sams argued in this direction and was in favour of attributing some early plays of unknown or doubtful authorship definitively to Shakespeare. Unfortunately the tools that were available to him were roughly the same that literary critics had used for centuries and there was an undeniable susceptibility to errors such as naming Shakespeare as author of the anonymous play *The Tragedy of Locrine* (see Sams 133-136).

It was only in the last few years, that new tools in computerised stylometry have been developed, and, especially with the availability of the program R Stylo from around 2013 (Eder, Kestemont, Rybicki), the authorship of the Marlowe corpus and early anonymous and apocryphal plays of English Renaissance drama could be assailed. The very latest approaches and methods of non-traditional stylometry have been applied and published in a number of cases (see literature links on my homepage *Shakespeare Statistics* <http://www.shak-stat.engsem.uni-hannover.de/>) (Ilsemann 2018, 2021). However, apart from in a few cases, I am not aware of any reaction. This may be due to the fact that the results deviate very strongly from what is considered to be traditional cultural knowledge, laid down for example in Wikipedia. With regard to the Marlowe corpus, the following diagram provides information:

Tamburlaine 1 Tamburlaine 2	
their style features	
can be found in (+)	are absent in (-)
anon. *The Tragedy of Locrine*	*Dido, Queen of Carthage*
Peele. *The Battle of Alcazar*	*The Jew of Malta*
Peele. *David and Bethsabe*	*The Massacre at Paris*
Kyd. *Cornelia*	*Edward II*
	Dr. Faustus (A) – 1604
	Dr. Faustus (B) – 1616

Figure 1 Stylistic equivalents of *Tamburlaine 1* and *Tamburlaine 2*

The two *Tamburlaine* parts are considered an important prelude in the development of the playwright Christopher Marlowe. That he actually wrote these plays himself is vouched for by historical evidence. The arrest of Marlowe in May 1593 was preceded by an event known as the Dutch Church Libel, which included the placement of a 53-line text on the wall of the Dutch cemetery in Broad Street in London. In bumpy verses of an iambic pentameter, Dutch immigrants were threatened if they did not leave the country. The poster was signed by Tamburlaine (see Carlo D. 2018). [1] The rest of the Marlowe corpus, it must now be said, has a completely uncertain attribution history. Marlowe's death on 30 May 1593 was a welcome occasion for printers and booksellers to make a profit with quarto prints attributed to him. The renowned Marlowe experts Kirk Melnikoff and Roslyn L. Knutson could only agree in their seminal work and included a chapter by Adam G. Hooks with the title Making Marlowe, written with the tenor that when Marlowe died "he was a corpse without a corpus." (Hooks 100)

Another complex of topics with far-reaching consequences, which has only opened up as a result of new computerised analyses, concerns William Shakespeare. When his former actor colleagues Heminges and Condell compiled the first Folio edition of the complete works in 1623, they criticised in the preface the many erroneous and incomplete printings of Shakespeare's plays, which were now "perfect of their limbes; and all the rest, absolute in their numbers as he conceived them." [2]

This, in some way, has a ring of self-amputation and of course it opened the door to the question as to how someone could so suddenly enter the stage with magnificent plays without working the least time as an apprentice. Even worse, in the Romantic period, not only was Shakespeare's genius celebrated, but at the same time more and more critics discovered the discrepancies between biography and works, and Shakespeare became the front man of the real author, who could not reveal himself for reasons of status or as an exile.

In fact, it is only modern non-traditional stylometry that is able to trace the development of the actor and young playwright and expand the number of his plays. Simple plots, folk-tale themes and romance-like elements such as in *Fair Em, Mucedorus, Edmund Ironside, A Knack to Know a Knave, The Life and Death of Jack Straw* and *George a Greene the Pinner of Wakefield* all showed stylistic similarities to, if not identity with, Shakespeare's work, some of which could only be determined using the *General Imposters* method (GI) and the complex Ružička distance, (Kestemont et al. 2016) which in line with *Rolling Delta* and *Rolling Classify*, also revealed Shakespearean parts in the remaining Marlowe corpus.

Kestemont et al. reported in 2016 on the role of nearest neighbours in determining the authorship of anonymous texts, and of the metrics used "to calculate the distances between vector representations of texts in a higher-dimensional space." In their study

---

1   It was not possible to decipher the author of this blog who abbreviated his name as Carlo D. For more details see Works Cited.

2   For the First Folio text, please see "Preface to *The First Folio* (1623)" – *Shakespeare Online*, http://www.shakespeare-online.com/biography/firstfolio.html [accessed 31 March 2022]

of the 'War Commentaries' by Julius Caesar (*Corpus Caesarianum*) they employed and evaluated the performance of the novel Ružička minmax metric and came to the conclusion: "Comparative evaluations across a variety of benchmark corpora show that this metric yields better, as well as more consistent results than previously used metrics." (246) This does not mean that the whole range of metrics has become unimportant. But the results of Burrows's delta (including Rolling Delta) and the highly acclaimed principal component analysis (PCA), for example, can now be checked with the General Imposters Method (GI) and its Ružička distance, supplemented by an optimization table with 'low' and 'high' values. Jan Rybicki explained in his script: "A score for a candidate which is equal or inferior to low [column B of Table 1] means that this candidate is NOT the author. A score for a candidate which is equal or superior to high [column C of Table 1] means that this candidate IS the author." 3

In 2018 GI had already been explained in detail in Eder's blog "Authorship verification with the package 'stylo'," published on the internet by the Computational Stylistics Group. The same webpage contains more information about cross-validations and performance measures. Those plays which are stylistically linked with the two *Tamburlaines* and the remaining plays of the generally acknowledged Marlowe corpus were checked against the reference plays, underlined in the passage below.

apo_locrine.txt, greene_friarbb.txt, greene_orlando.txt, kyd_soliman.txt, kyd_spanpure.txt, m0_fausta.txt, m1_bfaust.txt, m2_dido.txt, m3_edw2.txt, m4_jewmalta.txt, m5_massacre.txt, mar_tamburlain1.txt, mar_tamburlaine2.txt, mkyd_mscornelia.txt, mpeele_alcazar.txt, mpeele_davbeth.txt, nashe_summer.txt, peele_arraignment.txt, peele_oldwives.txt, shak_12thnight.txt, shak_hamlet.txt.

The tests comprised words (mf1w), word bigrams (mf2w), character bigrams (mf2c) and character trigrams (mf3c) which means that each of the checked texts was evaluated four times. Scores above the grey area are highlighted in bold white letters and a black background.

Table 1 Assignments with Ružička metrics

	A	B	C	D	E	F	G	H	I	J
1				assignments with Ruzicka metrics						variables
2	**plays**	low	high	**greene**	**kyd**	**mar**	**nashe**	**peele**	**shak**	
3	m0_fausta	0	0.72	0.08	0.42	0	0.06	0	**0.9**	mf1w
4	m0_fausta	0.12	0.79	0.01	0.34	0	0.02	0	**0.99**	mf2w
5	m0_fausta	0.18	0.81	0.08	0.49	0	**0.91**	0	0.52	mf2c
6	m0_fausta	0.23	0.66	0.02	0.35	0	**0.74**	0	0.65	mf3c
7				**greene**	**kyd**	**mar**	**nashe**	**peele**	**shak**	
8	m1_bfaust	0	0.72	0.16	0.35	0	0.17	0.02	**0.97**	mf1w
9	m1_bfaust	0.2	0.74	0.23	0.49	0	0.07	0	**0.99**	mf2w

3   Jan Rybicki's script makes use of cross-validations and performance measures, has not been published yet and was attached to a private email in February 2022.

				greene	kyd	mar	nashe	peele	shak	
10	m1_bfaust	0.17	0.81	0.12	0.42	0	**0.85**	0	0.67	mf2c
11	m1_bfaust	0.23	0.62	0.1	0.2	0	0.44	0	**0.98**	mf3c
12				greene	kyd	mar	nashe	peele	shak	
13	m2_dido	0	0.74	0.45	**0.91**	0.06	0	0	0.28	mf1w
14	m2_dido	0.05	0.85	0.29	**1**	0.17	0	0	0.31	mf2w
15	m2_dido	0.24	0.75	0.43	0.59	0	0.55	0.05	0.23	mf2c
16	m2_dido	0.3	0.61	0.25	**0.93**	0.01	0.12	0	0.48	mf3c
17				greene	kyd	mar	nashe	peele	shak	
18	m3_edw2	0	0.79	0.08	0.73	0.05	0	0	0.72	mf1w
19	m3_edw2	0.08	0.82	0.1	0.57	0.05	0	0	**0.99**	mf2w
20	m3_edw2	0.13	0.86	0.09	0.59	0	0.03	0	**0.9**	mf2c
21	m3_edw2	0.22	0.62	0.09	0.58	0	0.03	0	**1**	mf3c
22				greene	kyd	mar	nashe	peele	shak	
23	m4_jewmalta	0	0.8	0.01	0.56	0	0.01	0.01	**0.88**	mf1w
24	m4_jewmalta	0.34	0.65	0	0.44	0.01	0.02	0	**0.9**	mf2w
25	m4_jewmalta	0.18	0.81	0	0.53	0	0.35	0.04	**0.72**	mf2c
26	m4_jewmalta	0.05	0.79	0	0.29	0	0.07	0	**0.99**	mf3c
27				greene	kyd	mar	nashe	peele	shak	
28	m5_massacre	0	0.8	0.09	**0.82**	0.12	0	0	0.73	mf1w
29	m5_massacre	0.16	0.77	0.05	0.64	0.24	0	0	**0.9**	mf2w
30	m5_massacre	0	0.89	0.24	0.81	0.01	0.02	0.02	0.68	mf2c
31	m5_massacre	0.2	0.62	0.14	**0.66**	0.11	0	0	**0.92**	mf3c
32				greene	kyd	mar	nashe	peele	shak	
33	apo_locrine	0	0.79	0.04	0.3	**0.95**	0	0	0.09	mf1w
34	apo_locrine	0.19	0.71	0	0.14	**1**	0	0	0.13	mf2w
35	apo_locrine	0	0.78	0.14	0.16	**1**	0.1	0	0	mf2c
36	apo_locrine	0.23	0.64	0.04	0.23	**1**	0.1	0	0.02	mf3c
37				greene	kyd	mar	nashe	peele	shak	
38	mkyd_mscornelia	0	0.71	0	0.45	**0.93**	0.01	0	0.09	mf1w
39	mkyd_mscornelia	0.09	0.8	0	0.57	**0.92**	0.08	0	0.13	mf2w
40	mkyd_mscornelia	0	0.84	0.02	0.6	**0.84**	0.24	0	0	mf2c
41	mkyd_mscornelia	0.22	0.64	0	0.52	**0.91**	0.17	0	0.01	mf3c
42				greene	kyd	mar	nashe	peele	shak	
43	mpeele_alcazar	0	0.81	0.17	0.16	**1**	0	0	0.04	mf1w
44	mpeele_alcazar	0.17	0.82	0.03	0.37	**1**	0	0	0.05	mf2w
45	mpeele_alcazar	0.17	0.81	0.26	0.17	**1**	0.01	0.01	0	mf2c
46	mpeele_alcazar	0.16	0.65	0.08	0.33	**1**	0	0.01	0.14	mf3c

47				greene	kyd	mar	nashe	peele	shak	
48	mpeele_davbeth	0	0.76	0.14	0.22	**1**	0	0	0.05	mf1w
49	mpeele_davbeth	0.15	0.84	0.02	0.28	**1**	0	0	0.13	mf2w
50	mpeele_davbeth	0	0.84	0.13	0.16	**0.98**	0.01	0.22	0.16	mf2c
51	mpeele_davbeth	0.01	0.82	0.03	0.2	**1**	0	0	0.07	mf3c

In lines 3 to 31, the author column *shak* marks those assignments that indicate stylistic identity with Shakespeare's work. In lines 33 to 51, the author column *mar* shows those misattributed plays that actually originate from Marlowe's pen.

The consequences are immense. In a considerable part of the Marlowe-Shakespeare secondary literature, renowned literary critics deal with the influence of Marlowe's plays on Shakespeare's work, without even suspecting that in reality Shakespeare had always been compared with Shakespeare. The new Oxford Shakespeare edition of 2016 took the biscuit, officially declaring Marlowe to be the co-author of the three parts of *Henry VI*. The authorship companion that followed in 2017 made it clear that the stylistic proximity of *Edward II* to *Henry VI* was the reason for this decision. *Edward II* had been unquestioningly adopted according to tradition as a history play by Marlowe. However, all the research with R Stylo proves that the text was written by Thomas Kyd and William Shakespeare (see also attribution figures in lines 18 to 21).

This brief paper is nothing else but an exposé. With all the figures and tables that prove and undergird the aforementioned positions hundreds of pages could be filled. What is actually necessary is a wider phalanx of approaches, supported by many more researchers who are prepared and willing to combine traditional literary research with the results of modern non-traditional stylometry. May be this paper can offer an incentive to tackle literary traditions that have become questionable.

Works Cited

Clark, Arthur Melville. "Thomas Heywood as a Critic," *Modern Language Notes*, Apr., 1922, Vol. 37, No. 4 (Apr., 1922), 217-223, The Johns Hopkins University Press, <https://www.jstor.org/stable/2914670D> [accessed 29 March 2022].

Carlo, "Framing Marlowe. The Dutch Church Libel," *The Marlowe-Shakespeare Connection*, August 2008, <marlowe-shakespeare.blogspot.com/2008/08/framing-marlowe-dutch-church-libel.html> [accessed 30.09.2019].

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

Eder, Maciej. "Authorship verification with the package 'stylo'," Blog of the Computational Stylistics Group, May 30, 2018, https://computationalstylistics.github.io/blog/ [accessed 29 March 2022].

Hooks, Adam G., "Making Marlowe," in *Christopher Marlowe, Theatrical Commerce, And The Book Trade,* ed. by Kirk Melnikoff and Roslyn L. Knutson, Cambridge: University Press, 2018. 100.

Ilsemann, Hartmut , "A brief supplement to 'The Marlowe Corpus Revisited' and *Phantom Marlowe;*" *Digital Scholarship in the Humanities*, Advance publication (14 October 2021), https://doi.org/10.1093/llc/fqab078.

Ilsemann, Hartmut , "The Marlowe Corpus Revisited," *Digital Scholarship in the Humanities*, 36.2 (June 2021): 333-360.

Ilsemann, Hartmut, "Christopher Marlowe: Hype and Hoax", *Digital Scholarship in the Humanities*, 33.4 (2018): 788–820. https://doi.org/10.1093/llc/fqy001

Ilsemann, Hartmut, *Phantom Marlowe: Paradigmenwechsel in Autorschaftsbestimmungen des englischen Renaissancedramas*, Düren: Shaker, 2020, ISBN 978-3-8440-7412-3

Kestemont, M., J. Stover, M. Koppel, F. Karsdorp, W. Daelemans, W. , "Authorship Verification with the Ruzicka Metric. " in *Digital Humanities 2016: Conference Abstracts*. Jagiellonian University & Pedagogical University, Kraków, 2016. 246-249.

Preface to *The First Folio* (1623) – Shakespeare Online. < http://www.shakespeare-online.com/biography/firstfolio.html> [accessed 31 March 2022].

Sams, Eric. *The Real Shakespeare: Retrieving the Early Years, 1564-1594*, New Haven, London: 1995. 163-166.

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