

Who wrote Shakespeare?
Paradigmatic Changes in Authorship Attributions

by

Hartmut Ilsemann

(Leibniz Universität Hannover)

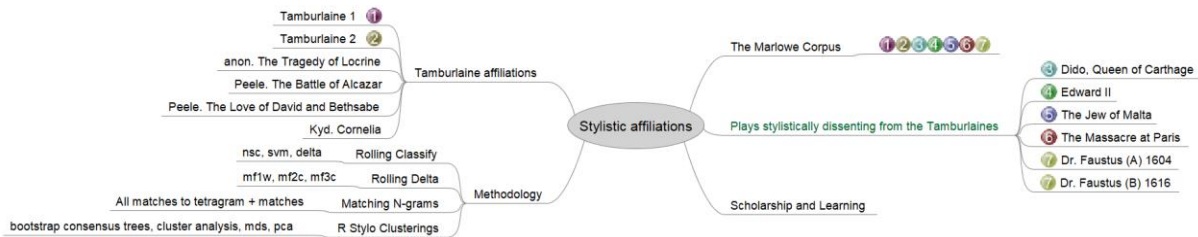


Fig. 1: Mind map of stylistic affiliations and methodological approaches (Source: Hartmut Ilsemann 2019, Fig. 1, generated with FreeMind)

In this paper, Hartmut Ilsemann argues against the authorship attribution put forward by the editors of the New Oxford Shakespeare edition (2016) who claim that Christopher Marlowe is a co-author of Shakespeare's *Henry VI* trilogy. This authorship attribution is primarily the result of the work of John Burrows and Hugh Craig who as members of the "attribution board" advised Oxford University Press on the matter of authorship. Their research is based on computational textual analysis using Delta, a stylometric tool introduced by Burrows in 2002. Ilsemann, on the other hand, uses the more recent tool Rolling Delta which is part of R Stylo.

R stylo, a suite of stylometric tools which incorporates a combination of powerful procedures and functions, was used to analyse the canon plays of Christopher Marlowe, who is generally regarded as the author of the two parts of the *Tamburlaines* as well as *The Tragical History of Doctor Faustus*, *The Jew of Malta* and *Edward II*. *Dido, Queen of Carthage* is counted among his works, either in part or in toto. A fragment with the title *The Massacre at Paris* also carries his name which has also been associated with poems and with translations of Ovid. *Hero and Leander* was left unfinished and Chapman completed the narrative verses. Furthermore it is said that Renaissance drama owes him some extraordinary characters and many examples of excellent blank verse. According to Thomas Dabbs and other scholars little was said or even known about Marlowe until the early 19th century, when this literary figure was rediscovered and historicised. People began to cherish dramaturgical aspects and Marlowe gained a place next to Shakespeare, Jonson, Beaumont and Fletcher. It is time to check the existing authorship ascriptions with modern non-traditional stylometry tools like R Stylo, in a situation where literary history is still perceived by men of letters who depend on what had been passed down from one generation to the next, including the attribution of plays to their authors. Stylometric analysis with R stylo is based on the authors' use of linguistic markers. These are in particular the highly frequent function words, such as "to" or "and", as well as character- and word n-grams, i.e. strings of an n-number of consecutive words and letters, such as "have been" or "and the". Authors tend to be unaware of how they use these types of words and lexical strings and they have proved to be characteristic of an author's style.

When Burrows introduced Delta as a measure of stylistic difference in 2002, he went far beyond of what traditional stylometry could do. Now it was not any more the comparison of two candidates at a time, but a new relation between the target text and a whole sequence of reference texts by various authors from which the best suitable could be drawn. However, Delta, being derived from absolute z-score differences and their mean, referred to whole texts. The reference texts had to be single-authored and well-attributed, a sine qua non for any correct attribution. What Delta could not do was to find collaborations in a play, unless the target text was parted and tested accordingly. Rolling Delta

overcame this shortcoming by dividing the texts into windows of a particular size. These windows (mostly 5000 words) roll through the texts with an overlap (250 words) and each time a measurement is taken. The huge number of measurements then goes into a spreadsheet and from each window the three lowest deltas are highlighted. The reference plays that do not have a single highlighted window are eliminated, leaving behind the best suited reference windows, which, transposed by 90 °, make up the attribution table. With regard to the Marlowe corpus it became clear that this was stylistically totally inhomogeneous. The style of *Tamburlaine 1* and *2* was not present in the remaining corpus. Instead it could be found in the anonymous play *Lochrine*, in Peele's *The Battle of Alcazar* and *The Love of David and Bethsabe*, and in Kyd's *Cornelia*. These findings were underpinned by results from Rolling Classify, matching unique n-grams between the plays, and graphics like multidimensional scaling, principal component analysis, cluster analysis, and bootstrap consensus trees. When *3 Henry VI* was analysed with Rolling Delta it became possible to attribute one half to Shakespeare and another half to Kyd. Craig and Burrows who had only one delta value to denote a whole text presented a list starting with *Edward II* and *The Spanish Tragedy* (Kyd) as far as the non-Shakespearian text was concerned. Dealing with the Shakespearian part the list comprised *Edward II* and *Richard III*. Their mistake was to accept *Edward II* as a play by Marlowe, relying erroneously on the teachings of traditional scholarship and learning.

Reference

Hartmut Ilsemann (2019). "The Marlowe Corpus Revisited." Submitted to *Digital Scholarship in the Humanities*, 1-40.