# Summary

In earlier articles, we saw that unique formal 3-grams and unique formal 4-grams are the most reliable for authorship attribution by the method I used. In this article I use the method with just those N-grams, to test authorship attributions for *Arden of Faversham* and the so-called extended Kyd canon.

We see below that the method accepts some, but not all, of the extended Kyd canon attributions. It supports the attribution of the whole of *Arden of Faversham* to Kyd. It also supports claims of the influence of Kyd on the young Shakespeare. However, for the reasons I give, this should not be taken as a rejection of the attribution of scenes 4-9 of *Arden* to Shakespeare.

# The Extended Kyd Canon

For these new tests I have made a further improvement. When the method rejects an attribution, then the winning author’s name is now also shown in the red cell, to save having to look it up from the raw data spreadsheet. To avoid clutter, I have not given the runner-up’s name in a green cell, but that information may sometimes be useful, and it can be looked up from the raw data.

I used the 86 plays used before, and added the extended Kyd canon, consisting of the following plays:

* *The Spanish Tragedy (excluding the Additions)*
* *Soliman and Perseda*
* *Cornelia*
* *Fair Em*
* *Arden of Faversham*
* *Edward III (excluding the Countess scenes and excluding 4.4)*
* *Henry VI, Part 1 (excluding Act 1 and excluding 2.4, 4.2 and 4.5)*
* *King Leir*

For this test, I added Kyd as an author and assigned all the above plays to him. The raw data and summary of the results are here:



We see that the first five plays listed above are all accepted by the method as being by Kyd, including the whole *Arden of Faversham* play. The other three are assigned to Marlowe, except that 4-grams assign the *Edward III* scenes to Kyd, by a strong margin.

The raw data reveals Kyd to be the runner-up for the *1 Henry VI* scenes and the runner-up using 3-grams for *Edward III*. However, it is Shakespeare, not Kyd, who is the runner-up behind Marlowe for *King Leir*, possibly because of matches with *King Lear* (though this needs to be checked).

Introducing Kyd into the mix causes changes to the Marlowe attributions. *The Jew of Malta* now shifts to Kyd, as does *Edward II* using 4-grams, albeit weakly. It is noticeable how, excepting the *Tamburlaine* plays – which are invulnerable because of the way they help each other with plentiful matches – Kyd and Marlowe appear to be almost interchangeable. They were room-mates, and Kyd was tortured because some of Marlowe’s words were mistakenly thought to be his. We will never know how many words in each man’s plays came from the other’s pen.

A more interesting change occurs to the Shakespeare attributions. Observe how some early plays shift attribution to Kyd, while the middle and late plays are unaffected. Among the early plays, only *The Comedy of Errors* stays with Shakespeare, but only weakly, and the runner-up is Kyd.

If we do not wish to attribute these early Shakespeare plays to Kyd, the results here are nevertheless good evidence of Kyd’s influence on the young Shakespeare, an influence that faded away within a few years of Kyd’s death. Now, as far as our method is concerned, Shakespeare is defined by *all* his plays listed; yet, if the middle and late plays do not contain much Kyd influence then the method will not detect much Kyd influence in Shakespeare as a whole. Therefore, when it finds Kyd phrases in the early Shakespeare plays, it will mistake them for Kyd’s writing. That seems to be the explanation why these early Shakespeare plays get attributed to Kyd. In turn, this reinforces the point that changes in an author’s style over time can cause some of his work not to be recognised as his by purely computational methods.

Hartmut Ilsemann has published two articles[[1]](#footnote-1) this year in which he has used Delta to relieve Marlowe of all his plays except *Tamburlaine* but assigned him *Locrine* and *Cornelia* in return. Perhaps this also testifies to the difficulty that computational methods have in distinguishing Marlowe from Kyd.

These results provide support for at least half of the extended Kyd canon. They also provide support for the New Oxford Shakespeare claim that Marlowe co-authored *1 Henry VI*, as well as Thomas Merriam’s claim that *King John* and *Henry V* were co-authored, albeit that the method identifies Kyd rather than Marlowe as the co-author.

# Scenes 4-9 of *Arden of Faversham*

I reverted to the 86 plays – that is, I removed the extended Kyd canon – and just added one division consisting of scenes 4-9 (or 3.1 to 3.6) of *Arden of Faversham*, assigning it to Shakespeare.



The method rejects the attribution, assigning the scenes to Marlowe (it did not have the option of assigning them to Kyd as he was not one of the authors available to it).

A similar test, this time replacing the *Arden* scenes 4-9 with the Countess scenes from *Edward III*, also led to the same outcome, i.e. the Shakespeare attribution was rejected.

Having considered why this happened, I am conscious that the method has not been shown to be reliable for small portions of plays, such as a handful of scenes. All tests until now have been on whole plays, and it would be hasty to assume that the method works with the same accuracy for parts of plays.

I need to study this further, to see why the method gives the results it gives for these scenes, perhaps also experimenting with co-authored plays such as *Henry VIII* and *The Two Noble Kinsmen*. Until then, it would be premature to say that this method has rejected the attribution of scenes 4-9 of *Arden* to Shakespeare.

# The Weighting Formula Revisited

I will need to look again at the weighting formula. Recall that our weighting formula is:

Where *T* is the total number of matches with author *A* for our play *P*, and *S* is defined as:

This formula was adapted from the one I used for the data I published in October 2017 and June 2018, for maximal and formal N-grams respectively. That earlier formula divided the number of matches (*T*) between two plays by the sum of the number of words in those two plays. I got the idea from Martin Mueller’s old blog post ‘Authors are Trumps’, as I acknowledged at the time.

MacDonald P. Jackson observed some months ago that my weighting formula is unsatisfactory if one play happens to be much longer than the other. The problem is much more acute when, as here, we are adding not the number of words in two plays, but the number of words in a play and the number of words in an author’s entire canon. Usually, the latter number will overwhelm the former. The result is not much different than we would get just by dividing by the number of words in the author’s canon alone.

It may be better to divide by the *product* of the number of words in the play and the number of words in the author’s canon, rather than to divide by their *sum*. That way, each play or division would exercise an influence on the weight exactly in proportion to its size, without its influence being overwhelmed by the size of the author’s canon. Whether or not this improves the results needs to be established by some experiments.

# A Few More Caveats

We must remember that any spreadsheet with the red/green/amber cells is a simplification of its corresponding raw data spreadsheet. That raw data spreadsheet is itself a simplification of the full set of data I published in June 2018, since it replaces several numbers of N-gram matches with an author’s plays by just one number for that author. Inevitably, some knowledge may be lost by the simplification. Neither a green nor a red cell excludes the possibility of co-authorship. If the winning or losing margin is small, it may be a clue that the play is co-authored, and the raw data may enable us to spot the likely co-author as the runner-up. We need to remember this when dealing with plays such as *Arden of Faversham*, *Edward III* or the *Henry VI* trilogy. The method will always give just one signal – red, green or amber – but the underlying raw data may also hold important clues not conveyed by the signal.

It is perfectly possible for a play to be assigned to one author, with a satisfactory winning margin, but the possibility may remain that there is some division of the play into two portions, assigned to different authors, with each attribution receiving an even more satisfactory winning margin. A successful result for one attribution does not automatically rule out every attribution that is incompatible with it. Life is not that kind. That is not a problem unique to this method; on the contrary, it is common to all methods, which is why scholars should be wary of claiming certainty for any attributions they make using any method. We know that it is possible for different methods to confirm the same attribution; in principle, it is also possible for the same method to confirm different and incompatible attributions, if we try enough tests.

Pervez Rizvi

October 2018

# Postscript

I investigated the failure of the method to attribute the Countess scenes to Shakespeare. The reason is simply the paucity of data.

For the Countess scenes, the number of unique formal 3-grams matches with Marlowe is 9, whereas with Shakespeare it is 50. The Shakespeare canon is much larger, so when we apply the weighting formula, Marlowe wins the attribution rather than Shakespeare. However, the result is volatile: if the Marlowe count had been 8 rather than 9, then Shakespeare would have won by a small margin.

It would be easy enough to find a weighting formula that enabled Shakespeare to win this attribution, but that would be likely to cause problems elsewhere by making other attributions incorrect. The real problem is that the numbers – in this case, just 9 unique matches with the entire Marlowe canon – are too small to use safely. It would be irresponsible to base any attributions on such exiguous data.

This experiment with the Countess scenes provides a further insight into the results we saw earlier. In the first article, “Which N-grams are the Best?”, we found that 4-grams and 5-grams did well using all matches, whereas 3-grams and 4-grams did well when using unique matches. The reason for the poor performance of all 3-grams was that the very common phrases that everyone used were a dominating influence, compromising the method’s ability to distinguish between the authors accurately. Because 5-grams are less common, even when they are not unique, the 5-grams test was able to do well. When I switched to using unique N-grams, the common phrases that everyone used were excluded, allowing 3-grams to do well. However, the paucity of unique 5-grams caused the test using them to do badly. 4-grams, being a compromise between the extremes of too few results and too many results consisting of common phrases, did well in both tests.

This method, if it is useful at all, is useful only for whole plays or substantial portions of plays, since it is only for them that we get results in the happy middle between the two extremes. Scenes 4-8 of *Arden of Faversham* are almost exactly half the size of the Countess scenes in *Edward III*. They are therefore even less suitable than the Countess scenes for this method.

Pervez Rizvi

November 2018

1. ‘Christopher Marlowe: Hype and Hoax’ and ‘Forensic Stylometry’ in *Digital Scholarship in the Humanities*. [↑](#footnote-ref-1)