

Hugh Craig

reviews

The Stylometric Analysis of
Faustus, from the German of Goethe

THE FOLLOWING is a brief review of the statistical section (pp. 312-30) of Frederick Burwick and James C. McKusick's edition of an anonymously published translation of a play by Goethe, *Faustus: From the German of Goethe: Translated by Samuel Taylor Coleridge* (Oxford: Clarendon Press, 2007).

The electronic corpus used for the Burwick and McKusick study is as follows: the anonymous 1821 *Faustus*, using the text prepared for the Oxford edition; five other translations of the play, by Germaine de Staël, George Soane, Daniel Boileau (this translation is called "the Boosey translation" in the stylometrics chapter), John Anster and Lord Francis Leveson-Gower, all again from the texts prepared for the edition; an adaptation of the Goethe play by George Soane, from an unspecified text; and two plays by Coleridge and two plays translated by him, all in texts from Chadwyck-Healy's Literature Online.

The editors draw on two sets of data derived from this corpus, the first the frequencies of words of two letters, three letters and so on, up to words of eight letters, and the second the frequencies of ten particular words (*he, in, now, of, shall, then, this, to, which, and your*) which they find are used at different rates in a Coleridge play and in a group of translations of *Faustus* by other writers.

They use the chi-squared test to determine whether the differences between the counts for the anonymous 1821 *Faustus* and various plays and groups of plays are significant or not. If the differences are significant, they take this fact to be evidence that the author of the comparator play is not the author of the 1821 *Faustus*. If the differences are not significant, they take this fact to be consistent with notion that the author of the comparator play is also the author of the 1821 *Faustus*. The chi-squared test and the authors' conclusions from its results are well within normal stylometric practice.

The first set of data, the distribution of words by number of letters, is an approach which (as the authors mention) goes back to Thomas Mendenhall and the 1880s. Currently it is not widely practised. As the authors say twice on p. 313 and once on p. 316, it has proved unreliable in tests following Mendenhall's early experiments, which showed that the profile of Bacon's work was quite unlike that of the Shakespeare plays, while Marlowe had a profile indistinguishable from Shakespeare's. C. B. Williams showed in 1975 that the test was confounded by differences in text type like verse versus prose,¹ and the Marlowe result suggests that the word length technique is what Ward E. Y. Elliott and Robert J. Valenza in a different context call a "Cinderella's slipper" method, i.e. one in which a number of candidates can

¹ C. B. Williams, "Mendenhall's Studies of Word-Length Distribution in the Works of Shakespeare and Bacon," *Biometrika* 62 (1975): 207-12.

give positive identifications.² Given the reservations it seems odd for Burwick and McKusick to put any weight on the findings. It seems contradictory, in fact, to note that “this kind of analysis is no longer considered definitive or even particularly reliable by stylometrists” and yet claim that “it is nevertheless possible to gain interesting and suggestive results by looking at this kind of data” (p. 316). One wonders whether the editors would have found the results “interesting and suggestive” or worthy of mention if they had been less favourable to the Coleridge case.

In the event the analysis shows that the 1821 *Faustus* fits the profile of one of the Coleridge plays, *Remorse*, but not those of the other three. The disputed play’s profile is also unlike those of the other five translations or the Soane adaptation. For a Coleridge attribution, this is a mixed result, since the closeness to *Remorse* is certainly consistent with Coleridge’s authorship of the anonymous translation, while its remoteness from the other three Coleridge plays is evidence against. The authors explain the discrepancy by referring to the proximity of the date of *Remorse* and the 1821 version and their likeness in genre and subject matter. The first of these we can discard, since there is another Coleridge play in the set (*Zapolya*) closer in date (p. 316). The second, similarity of genre and subject matter, is a two-edged sword: it helps explain why only one of the Coleridge plays might resemble the 1821 *Faustus*, but it also gives a possible explanation other than shared authorship why the plays might resemble each other in profiles of word length distribution.

In their text the authors make no secret of the fact that the 1821 *Faustus* proves to be unlike the other three Coleridge plays. They choose, however, to show graphs and tables only for the results that support the case for Coleridge as author of the disputed play, i.e. the comparisons between the 1821 *Faustus* and *Remorse*, and between the 1821 *Faustus* and the five other translations. This means that readers cannot judge just how different the other Coleridge plays are from the 1821 translation on this test.

The second data set is the frequencies of the ten function words in the various texts. These words were chosen from lists of words which discriminated between the Coleridge plays in the set and the translations by others, and had similar frequencies in the two early Coleridge plays, considered as one text, and in the two later Coleridge plays, also considered as one text. The profile of the ten words is very similar in the case of the 1821 translation and *Remorse*. The profiles of the disputed translation and the translations by others are quite unlike each other.

This time the comparison between the 1821 translation and the three other Coleridge plays is not mentioned. Perhaps the authors thought that the consistency within the Coleridge corpus generally for the set of ten words was evidence enough, but it would certainly have been useful confirmation if they could have shown that the 1821 translation is also close to the other three

² Ward E. Y. Elliott and Robert J. Valenza, "Glass Slippers and Seven-League Boots: C-Prompted Doubts About Ascribing *A Funeral Elegy* and *A Lover's Complaint* to Shakespeare," *Shakespeare Quarterly* 48 (1997): 177-207.

individually on the profile of these ten words. It would also have been instructive to compare counts for the words in a composite text of the four Coleridge plays against each of the translations in turn. This would follow the logic of the design of the keyword set, which the authors say was meant to be “a common set of keywords... that distinguishes Coleridge’s usage from all of the other candidates” (p. 322), “a set... that consistently reflects Coleridge’s vocabulary in all of his verse dramas” (p. 324).

The authors are careful not to claim any more than “strong likelihood” for Coleridge’s authorship from the tests. Generally the study does indeed seem indicative of Coleridge’s authorship rather than conclusive. The candidates beyond Coleridge in the study do not to my mind offer particularly strong competition (in that it seems unlikely that a writer who had produced one translation would then produce a second quite different one) and are not given the same opportunities as Coleridge to display resemblances to the disputed text (four separate plays by Coleridge are tested, but only a single work by each of the other five). It might be argued that these are the only likely candidates, and that a translation of the same work gives each such a good opportunity to resemble the 1821 *Faustus* that further testing is unnecessary. What is really needed, though, is a preliminary assessment of the accuracy of the methods in assigning comparable works (translations of early nineteenth-century plays, if enough can be found, or texts as close as possible to this text type) to their known authors, and away from other authors, when they are treated as anonymous pieces. This would be especially valuable if the authors mean to rehabilitate the word-length profile method. One result which does seem substantial and important is that the anonymous translation is quite different on two tests from six other versions from the same period. Anyone who suspected that the anonymous translation was a version of one of the other translations can be reassured that this is quite a different animal.

Optimists could take away from the study the idea that the anonymous translation is indistinguishable from Coleridge’s play *Remorse* and quite different from six translations and adaptations by other writers, on two separate measures. The pessimistic view would be that the target text is like one out of four of the Coleridge plays examined, and unlike a set of other texts, on two tests, one long since discredited, the other with a good pedigree, but not calibrated or evaluated for this sort of text.

The authors often invite replication of their study in the chapter and say that their use of the *Signature* software, available gratis from the University of Leeds, and their choice of Literature Online e-texts for the four Coleridge plays, makes this possible (pp. 315, 316, 327, 330). In fact replication is not feasible without e-texts of the anonymous translation and the other *Faustus* versions. There is no indication in the edition that these are available to the scholarly community generally. Nevertheless, the use of the *Signature* program does have the signal advantage that it is possible to study Burwick and McKusick’s methods in some detail. One other comment: the two procedures

drawn from *Signature* derive from Mendenhall's work on Shakespeare in the 1880s, as already mentioned, and the work of Frederick Mosteller and David L. Wallace on the *Federalist* papers, published in 1964. It is a little disheartening for those who continue to practise statistical attribution that the considerable effort that has gone into this kind of work in the last four and a half decades leaves no trace in the methods used in the Burwick and McKusick study or indeed in the version of *Signature* they employed.

How much does the statistical section of the edition add to the case for Coleridge as author of the translation? I think what we have here are more "straws in the wind" than anything like conclusive proof (which the authors, it must be admitted, do not claim in any case). Without fresh supporting work on a control group, to reverse failures in previous studies, the word length test is no more than a curiosity. The word frequency test is much more promising. To make it more than that requires something we are not given in the chapter, a thorough prior validation, which would establish whether the test can distinguish regularly between samples of Coleridge's dramatic writing and samples of comparable writing by others where authorship is known. As things stand, my view would be that the results of this test do add some support to the case for Coleridge, but of a minor order.