Neural Computation in Stylometry II: An Application to the Works of Shakespeare and Marlowe

THOMAS V. N. MERRIAM
Basingstoke, UK

ROBERT A. J. MATTHEWS
Oxford, UK

Abstract

Using principles set out in an earlier paper, a neural network was constructed to discriminate between the works of Shakespeare and his contemporary Christopher Marlowe. Once trained using works from the core canon of the two dramatists, the network successfully classified works to which it had not been previously exposed. In the light of these favourable results, we used the network to classify a number of anonymous works. Strong support emerged for Tucker Brooke's view that *The True Tragedy* is the Marlovian original of *Henry VI, Part 3*, the latter being the product of subsequent revisions by Shakespeare.

1. Introduction

In an earlier paper (Matthews and Merriam, 1993; henceforth paper I), we argued that pattern recognition techniques based on so-called neural computation constituted a useful addition to the stylometric toolkit. We showed that a multi-layer perceptron (MLP) neural network could be trained to discriminate between the works of Shakespeare and his contemporary John Fletcher with considerable reliability, and went on to use the MLP to investigate claims of collaborative authorship of certain plays.

In this paper, we extend the application of neural computation to an investigation of works by Shakespeare and perhaps his most famous literary contemporary, Christopher Marlowe (1564–93). As before, our aim has been to create an MLP capable of discriminating between the works of these two dramatists, and then use it to investigate works of contentious authorship.

Neural networks constitute an attractive new technique for stylometry, as they are known to be capable of dealing with both a considerable amount of noise and non-linear interactions within a data-set, both notorious problems in conventional stylometry. These considerable advantages do not, of course, imply that conclusions based on MLPs are infallible. Rather, by providing a new approach to studying stylometric characteristics, they provide a potentially valuable source of further evidential weight upon which scholars can draw.

We begin by briefly recapping on how stylometric MLPs may be designed and trained to perform this task; for more technical details, see paper I.

Correspondence: Dr Thomas Merriam, 35 Richmond Road, Basingstoke RG21 2NX.

2. Building a Shakespeare/Marlowe MLP

A stylometric MLP consists of a set of m input neurons, each one of which represents the numerical value of a stylometric characteristic ('discriminator') capable of distinguishing between the works of Shakespeare and Marlowe. Each of these inputs is then connected to a second layer of neurons, the so-called hidden layer, the strength of connection being dictated by trainable weights and biasses. The hidden layer is, in turn, connected to an output layer, consisting of just two neurons, corresponding to the two authors. Figure 1 shows a typical MLP topology.

To enable an MLP to discriminate between different authors, the values of the weights and biasses have to be trained. This is done by giving an MLP input values drawn from samples of the authors' work, and then the desired output. At first, the MLP does no better than chance at linking each set of inputs to the desired author. However, by iterative use of the back-propagation algorithm (Rumelhart and McClelland, 1986), the weights and biasses of the MLP converge to values that enable the network to classify inputs with progressive accuracy.

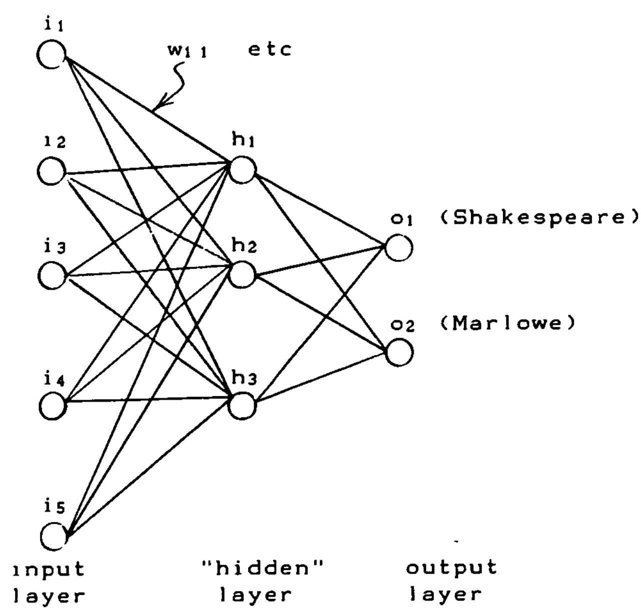

Fig. 1 A stylometric multi-layer perceptron

2.1 *Choice of Discriminants*

In setting up a stylometric MLP, the first task is to decide which stylometric measures will serve as suitable discriminators of authorship. The problem of discriminating between Marlowe and Shakespeare has already been studied in depth by one of us (Merriam, 1993). He found that the following five stylometric ratios have statistically highly significant discriminatory power:

no/T10, (of *x* and)/of, so/T10, (the *x* and)/the, with/T10

Here *x* represents any word and T-10 is Taylor's ten function words (but, by, for, no, not, so, that, the, to, with; see Taylor, 1987). Analysis by Merriam shows that these five tests separate Marlowe's *Tamburlaine the Great* from the Shakespeare core canon by an average of 3.6 standard deviations, and remain acceptably stable and strong discriminators over the Marlowe corpus.

By using these discriminators as inputs to a neural network, we can expect to arrive at a particularly powerful discriminator system which combines the power of the original individual discriminators with an ability to cope with noise and non-linear interactions among them.

2.2 *Training the MLP*

To train the network, fifty samples of Shakespeare's dramatic writing, consisting of acts of the same ten core canon plays used in paper I, were chosen. The plays, which span his career, are *Antony and Cleopatra*, *As You Like It*, *Henry IV*, *Part One*, *Henry V*, *Julius Caesar*, *Love's Labours Lost*, *Midsummer Night's Dream*, *Richard III*, *Twelfth Night*, and *The Winter's Tale*. For Marlowe the choice of plays was more restricted: *Tamburlaine I*, *Tamburlaine II*, and *Edward II*.

Our source for the Shakespeare First Folio plays was the machine-readable texts produced by the Oxford University Computing Service, while for the Marlowe plays we used those produced by Louis Ule. Despite the smaller canon size, it was still possible to extract fifty samples of 1,000 words for training purposes.

Once the five sets of 100 ratios were extracted for each discriminator, each set was normalized to give zero mean and unit standard deviation; this ensures that each discriminator contributes equally in the training process.

2.3 *Training Criteria*

Successful training of an MLP is typically gauged by so-called cross-validation. The MLP learns to perform its task using a proportion of the training set, and its success is then measured by using it to classify the part of the training set it has never seen before. Cross-validation helps combat the problem of 'overtraining', in which the MLP learns to classify the training set perfectly, but loses the power to generalize when exposed to new data (see, e.g. Hecht-Nielsen, 1989). After some experimentation, it emerged that we could reasonably expect to obtain cross-validation accuracies of 90% without running into overtraining problems.

Thus the first of our criteria for the completion of training was that the MLP be capable of classifying the cross-validation data with an accuracy of 90%.

The other criterion was set by the requirement that the MLP be unbiassed in its discrimination process; in other words, that it was no more likely to misclassify works of Marlowe as those of Shakespeare than it was to do the reverse. Thus we required that the MLP's misclassified data be approximately equally divided between the two classifications.

These criteria were then used to find a suitable size for the hidden layer. Too few hidden units fails to capture all the features in the data, while too many leads to a failure to generalize; in tests, we found that three hidden units were sufficient to give cross-validation results meeting our criteria.

We were then able to fix the topology for the stylometric MLP at five inputs, three hidden units, and two outputs.

The Merriam-based Shakespeare–Marlowe network (henceforth MERMAR) was found to successfully meet the training criteria after twenty or so presentations of the complete 100-vector training set MERMAR achieved a cross-validation accuracy of 90%, with the 10% misclassified being split into 6% Shakespeare classified as Marlowe, and 4% Marlowe classified as Shakespeare

2.4 *Testing and Performance Appraisal*

Having been trained, MERMAR was tested by being asked to classify undisputed works of Shakespeare and Marlowe that it had not seen during training. This constitutes a test of the power of the network to generalize to new data.

For the test, the network was asked to classify the remaining twenty-six entire plays of the Shakespeare First Folio, and the four remaining plays attributed to Marlowe: *Doctor Faustus* (1616), *Dido, Massacre at Paris*, and *The Jew of Malta*. Table 1 shows the signal strength from the two MERMAR outputs, corresponding to Marlowe and Shakespeare, for each of the thirty plays. The signal strengths are also shown in the form of a ratio, the Shakespearian Characteristics Measure (SCM); it is defined as:

$$\mathrm{SCM} = \Omega_S/(\Omega_S + \Omega_M) \qquad (1)$$

where Ω_S and Ω_M are the values of the outputs from the Shakespeare and Marlowe neurons, respectively Thus the SCM ranges from around unity for strongly Shakespearian classifications ($\Omega_S \gg \Omega_M$) to 0 for strongly Marlovian ones, and 0.5 for borderline classifications ($\Omega_S = \Omega_M$). The final column gives the corresponding verdict of MERMAR on authorship.

We see that MERMAR disagrees with conventional scholarship on the authorship of just two of the thirty plays: *The Jew of Malta* (given to Shakespeare by the MLP) and *Henry VI, Part 3* (given to Marlowe) This success rate of 93% compares well with the 90% rate expected on the basis of the cross-validation results.

The statistical significance of so close an agreement in these classification results can be judged by using the binomial distribution to calculate the probability $P(S)$ of obtaining at least S successes in T trials simply by

Table 1 Results for the rest of the Shakespeare and Marlowe canons

Play	Marlowe signal	Shakespeare signal	SCM	MERMAR verdict
SHAKESPEARE				
1H6	0 097	0.947	0 91	S
2H4	0.058	1.022	0 95	S
2H6	0 318	0 681	0 68	S
3H6	0.601	0 409	0.41	M[1]
ADO	−0 051	0 994	1.05	S
AWW	−0.003	0 984	1 00	S
CE	0 229	0 768	0 77	S
COR	0.102	0 923	0 90	S
CYM	−0 043	1 077	1 04	S
H8	0.025	0.965	0 98	S
HAM	0 006	0 941	0 99	S
JN	0.399	0.659	0 62	S
KL	−0.026	0.977	1 03	S
MAC	0.211	0.799	0 79	S
MM	0.153	0.844	0 85	S
MV	0 129	0 889	0 87	S
MWW	0 076	0.931	0 93	S
OTH	0 188	0 793	0 81	S
R2	0 095	0 955	0.91	S
ROM	0.110	0 930	0.89	S
TGV	−0 021	1 011	1.02	S
TIM	−0 001	1.039	1 00	S
TITUS	0 388	0 674	0 64	S
TMP	−0 030	1 043	1 03	S
TRO	−0 025	0 972	1 03	S
TS	−0 043	1 009	1 05	S
MARLOWE				
Dido	0 720	0.376	0 34	M
Faust	0.777	0.231	0 23	M
Jew Malta	−0 104	1.036	1 11	S[1]
Massac	0 850	0 125	0 13	M
ANONYMOUS				
EDW3	0 494	0 587	0 54	S
Conten	0 841	0 091	0 10	M
Tragedy	0 806	0 148	0 16	M

[1] Indicates prima facie disagreement with conventional verdict.

chance, given two equally likely outcomes. In our case, we have $T = 30$ and $S = 28$, so that $P(28) = 4.3 \times 10^{-7}$; the correct classification of twenty-eight entire plays by MERMAR is thus highly significant ($P \ll 0.001$). Table 1 is graphically displayed in Figure 2. The diagonal line is the locus of points for which the Shakespeare signal strength equals that of the Marlowe signal. Thus those points above the diagonal are classified as Shakespearian by MERMAR, and those below as Marlovian.

Although most plays lie well inside their respective regions, two lie relatively close to the Shakespearian–Marlowe borderline: *King John* and *Titus Andronicus*. Conventional scholarship provides grounds for doubting a strong Shakespearian signal for both of these plays. The former is related to an earlier two-part play, *The Troublesome Reign of John, King of England*, whose authorship is unknown. *Titus Andronicus* is closely related to an early quarto version, and its purely Shakespearian nature remains doubted by contemporary scholars (Taylor, 1987).

3 Analysis of the Problematic Plays

The substantive interest of Fig. 2 lies in the 'difficult cases': those plays whose authorship has apparently been wrongly attributed by MERMAR. There are two such plays: *The Jew of Malta* and *Henry VI, Part 3*.

It is, of course, possible that these misclassifications are simply a reflection of the ~10% classification error rate found during cross-validation of the MLP. However, it should be recalled that the cross-validation was carried out on data drawn from 1,000-word samples, whereas the play classification test was based on substantially larger, and thus less noisy, data. We might therefore well expect MERMAR to have >90% reliability when used on entire plays.

Certainly, the plays 'wrongly' classified by MERMAR have long been at the centre of doubts over their authorship on other, more subjective, grounds. Concern about the state in which the text of *The Jew of Malta* has reached us have led a number of scholars to exclude it from the Marlovian core canon; such exclusion certainly seems justified by the findings of MERMAR. Of all the plays tested *The Jew of Malta* shows the lowest Marlowe signal, placing it as an outlier in Fig. 2. As an Elizabethan play, it has a restricted affinity to *Arden of Faversham*, with concomitant possible attribution to Kyd.

Since collaboration cannot be ruled out, the problematic nature of *The Jew of Malta* may be due to

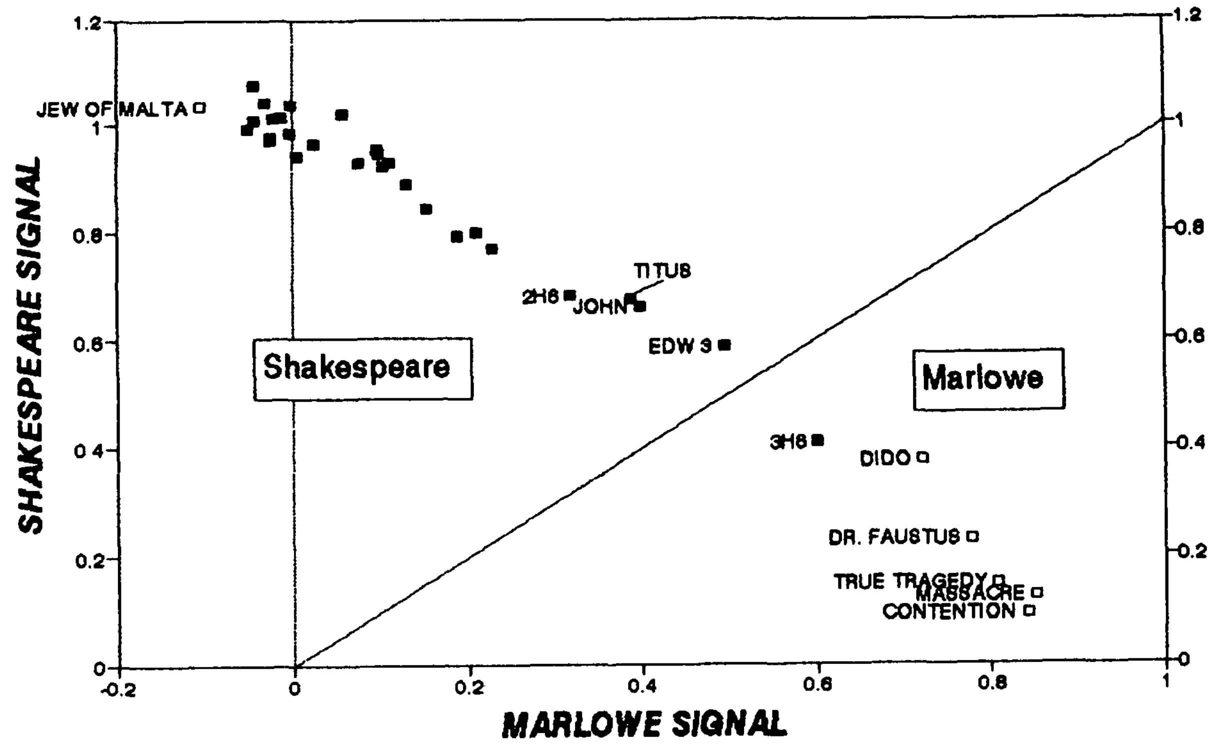

Fig. 2 Classifications of Shakespeare (closed squares) and Marlowe (open squares) plays not previously seen by MERMAR

mixed authorship. In effect, it may be argued that the neural network has, by placing the work in the Shakespearian category, simply declared against Marlowe's authorship.

Henry VI, Part 3 is ascribed by MERMAR to Marlowe, in contrast to the conventional Shakespearian attribution. However, Tucker Brooke (1912) has argued strongly for this play being a Shakespearian revision of a Marlovian original, *The True Tragedy of Richard Duke of York* (1595). This leads to the expectation that *Henry VI, Part 3* may reflect original Marlovian characteristics, despite Shakespeare's revisions; strong evidence supporting this interpretation is discussed below. Although not going as far as Tucker Brooke, contemporary scholars also continue to have doubts about the purely Shakespearian nature of *Henry VI, Part 3* (Taylor, 1987).

4. The Anonymous Works

Three anonymous plays were classified by the neural network: *Edward III*, *The First Part of the Contention* and *The True Tragedy*.

MERMAR's attribution of *Edward III* to Shakespeare agrees with many contemporary scholars, such as Taylor and others. However, we note that this play has the lowest SCM (0.54) of any play classified as Shakespearian by MERMAR. This lends support to Robertson's (1924) view that Marlowe had a considerable influence on the play, a view also partly supported by the findings of Merriam (1993), who gives stylometric arguments for Marlowe's authorship of scenes III.i and ii in this play.

It is interesting to speculate further that in the results for the three anonymous plays we are seeing Marlowe's influence in the creation of the history play in general, and the English history play in particular. Indeed, MERMAR may even be pointing to actual collaboration between Shakespeare and Marlowe. Certainly, of these three plays, neither *The Contention* or *The True Tragedy* are so late that co-authorship can be ruled out on chronological grounds.

Tucker Brooke considered both *The Contention* and the Octavo *The True Tragedy* to be Marlovian, views supported by MERMAR. In addition, as noted earlier, he argued that the latter was revised by Shakespeare to become the Folio *Henry VI, Part 3*. Until now, both support and criticism of this intriguing claim have relied primarily on subjective argument. MERMAR, however, gives us the opportunity to investigate the claim in quantitative detail.

5. The Case of *Henry VI Part 3*

Tucker Brooke's argument leads to the expectation that the Octavo *The True Tragedy* should show strong Marlovian characteristics when analysed by MERMAR, and that in the Folio *Henry VI, Part 3* these characteristics should be somewhat diluted by Shakespeare's alleged revisions.

This is just what MERMAR found, as Fig. 2 shows. Both plays lie in the Marlowe region, but while *The True Tragedy* lies deep within it (SCM = 0.16), *Henry VI, Part 3* is positioned much closer to the Shakespeare region, with an SCM 2.6× higher, clearly supporting Tucker Brooke's claim.

It is, however, possible to go further. On Tucker Brooke's argument, *Henry VI, Part 3* is the Marlovian

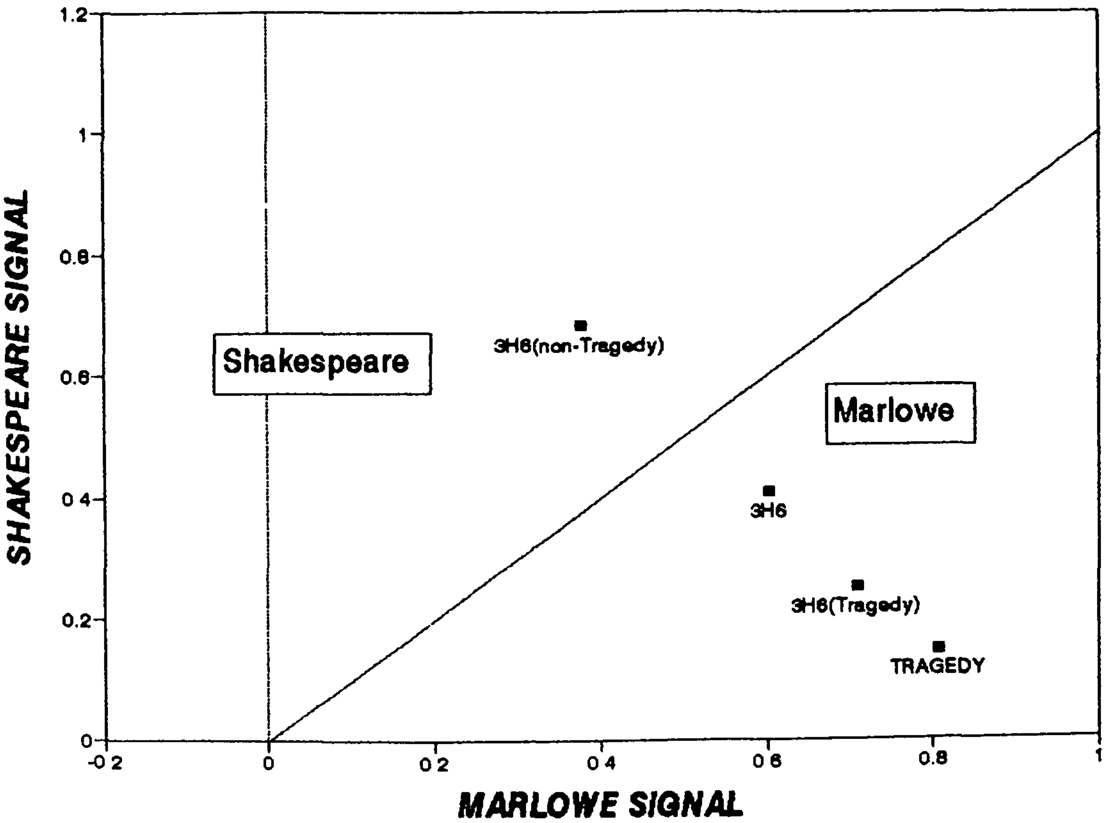

Fig. 3 MERMAR findings for Tucker Brooke's hypothesis

The True Tragedy with Shakespearian additions. Thus subtracting from *Henry VI, Part 3* those parts not also in *The True Tragedy* should leave behind a more strongly Marlovian version of *Henry VI, Part 3*—henceforth called *3H6* (Tragedy)—and a residue of Shakespearian additions which we call *3H6* (non-Tragedy). (Note: Taylor's decision to rename *Henry VI, Part 3* as *Richard Duke of York* while retaining the title *The True Tragedy of Richard Duke of York* for the Octavo can lead to confusion. We therefore eschew Taylor's usage.) This leads to the prediction that *The True Tragedy* and *3H6* (Tragedy) should have the lowest SCMs, followed by the extant version of *Henry VI, Part 3*, and finally *3H6* (non-Tragedy), which, being allegedly Shakespearian revisions, should have the highest SCM of all.

Table 2 shows the results of the MERMAR analysis, while Fig. 3 shows the relationship of the plotted points for the four elements. The data given in Table 2 and Fig. 3 are in the relationship predicted on the basis of Tucker Brooke's view that the Octavo *The True Tragedy* is a Marlovian original, revised by Shakespeare to become the Folio *Henry VI, Part 3*.

The Oxford Editors take a somewhat different view of the relationship between these two works, however. The former is considered to be a 'memorially reconstructed' version based on an edited, abridged and now lost Shakespearian original to which *Henry VI, Part 3* is more faithful.

It is not clear how this alternative theory can be tested by MERMAR, or indeed by any other stylometric technique; it is impossible to tell whether the result of the memorial reconstruction would be to increase or decrease the SCM of the original.

Nevertheless we can say that, at the very least, the results presented here suggest that Tucker Brooke's views deserve further consideration.

6. Conclusions

In this paper, we have used the technique of neural computation to develop an MLP capable of distinguishing between the works of Shakespeare and Marlowe. Cross-validation and testing on works by both dramatists previously unseen by the MLP showed that it is capable of providing results of considerable

Table 2

Test	*TT* (Octavo)	*3H6* (Tragedy)	*3H6* (Folio)	*3H6* (non-Tragedy)
M signal	0 806	0.710	0.601	0.378
S signal	0 148	0 253	0 409	0 682
SCM	0 16	0 26	0 41	0.64

reliability. When the MLP was applied to anonymous works *Edward III* emerged as essentially Shakespearian, while both *The Contention* and *The True Tragedy* emerged as strongly Marlovian in character.

Although the MLP's attribution of *Edward III* agrees with that of contemporary scholarship, its relatively weak Shakespearian character (SCM = 0.54) may well provide reason for pondering whether the argument for collaborative authorship can be dismissed.

In attributing both *The True Tragedy* and *Henry VI, Part 3* to Marlowe, the MLP supports Tucker Brooke's claim that the former is a work of Marlowe which was partly revised by Shakespeare to become the latter. Detailed analysis of the MLP's results confirm this.

We thus conclude that Tucker Brooke's arguments for the anonymous plays *The Contention* and *The True Tragedy* being by Marlowe, and that *Henry VI, Part 3* is a Shakespearian revision of a Marlovian original, all deserve renewed consideration.

On the broader issue of the use of neural networks in stylometry, we hope that the present paper and its predecessor will encourage others to go further. These introductory papers have shown how neural networks can be used to add evidential weight in binary questions of authorship for which a priori probabilities already exist. However, there is clearly scope for extending the use of neural networks to more complex, multinomial questions.

Acknowledgements

It is a pleasure to thank Professor David Bounds and Professor Chris Bishop of Aston University and Paul Gregory and Dr Les Ray of Recognition Research for their advice on neural computation issues. We also thank Louis Ule and Professor G. Richard Proudfoot for machine-readable texts of the Marlowe plays, and *Edward III*, respectively.

References

Hecht-Nielsen, R. (1989). *Neurocomputing.* Addison-Wesley, Reading.

Matthews, R. A. J and Merriam, T. V. N. (1993). Neural Computation in Stylometry I: An Application to the Works of Shakespeare and Fletcher. *Literary and Linguistic Computing*, 8.4: 203–9

Merriam, T V. N. (1993). Marlowe's Hand in *Edward III*, *Literary and Linguistic Computing*, 8.2: 59–72

Robertson, J M. (1924). *An Introduction to the Study of the Shakespeare Canon.* Routledge, London

Rumelhart, D E. and McClelland, J. L. (eds) (1986) *Parallel Distributed Processing (I).* MIT Press, Cambridge.

Taylor, G. (1987). The Canon and Chronology of Shakespeare's Plays. In *William Shakespeare. A Textual Companion* Clarendon Press, Oxford.

Tucker Brooke, C. F (1912). The Authorship of the Second and Third Parts of 'King Henry VI'. *Transactions of the Connecticut Academy of Arts and Sciences*, 47: 145–211.