Changes in style in authors with Alzheimer's disease

Graeme Hirst and Vanessa Wei Feng

(University of Toronto)

Abstract: Even in its very early stages, Alzheimer's disease leads to changes in language that can be detected by computational analysis. These changes may include a reduced, vaguer, and more abstract vocabulary, and reduced syntactic complexity. But do these changes affect an author's essential style? We experiment with a number of standard features for authorship attribution and authorship verification to see whether they recognize late works written by authors known to have had Alzheimer's disease as being by the same author as their earlier works. The authors whom we study are Iris Murdoch and Agatha Christie. Our control author (without Alzheimer's) is P.D. James. Our results were equivocal, as different frameworks yielded contrary results, but an SVM classifier was able to make age discriminations, or nearly so, for all three authors, thereby casting doubt on the underlying axiom that an author's essential style is invariant in the absence of cognitive decline.

Graeme Hirst is a Professor and Vanessa Wei Feng a PhD student in the Department of Computer Science, University of Toronto, Ontario, Canada. Email: gh@cs.toronto.edu, weifeng@cs.toronto.edu

1. Language changes in Alzheimer's disease

In the dementia associated with Alzheimer's disease, all cognitive abilities decline, including language, and even though there is considerable variation across individuals, there are regular patterns of linguistic deficits in Alzheimer's dementia that differ from those in other dementias and aphasias.[1] These changes may include a reduced, vaguer, and more abstract vocabulary, and reduced syntactic complexity, with less-frequent use of the passive voice, and they differ greatly in magnitude from the mild changes normally seen in healthy aging.[2] Thus a decline over time in an individual's richness of vocabulary and complexity of syntax that is greater than the normal changes of aging may be an indication of Alzheimer's disease.[3]

Le, Lancashire, Hirst, and Jokel studied changes over time in the language of three British novelists: Iris Murdoch, Agatha Christie, and P.D. James. Iris Murdoch is known to have died of Alzheimer's disease[4] and declined greatly while writing her last novel, *Jackson's Dilemma*. Agatha Christie, whose final novels were of low quali-

[1] Maxim and Bryan; Nicholas *et al.*

[2] Maxim and Bryan.

[3] Comparisons of these measures across individuals are not meaningful; what matters is change within the individual.

[4] Garrard.

ty, was suspected of having Alzheimer's disease but was never diagnosed.[5] P.D. James was still fully active, her health not in doubt, at the time of Le *et al.*'s study.

Le *et al.* found that in Murdoch's and Christie's late novels, their vocabulary was less rich, they repeated phrases more often, they used many more vague words and filler words, and overall they used fewer nouns and more verbs, whereas James showed no such changes. Murdoch and Christie also showed decreases in the use of the passive voice, consistent with Alzheimer's disease, and again James did not. However, Le *et al.* did not observe in Murdoch and Christie the more-general drop in syntactic complexity usually associated with Alzheimer's disease. But there are wide individual variations in the effects of the disease, and Le *et al.* concluded from their results that Agatha Christie probably did have Alzheimer's disease, and that her last three novels were written in a period of serious decline. Moreover, they showed that the early stages of the decline are visible in her novels of the preceding decade. Le *et al.* also found that Murdoch showed a significant decline and possible indications of problems even in her 50s, which were a troubled time for her,[6] with a recovery in her 60s before her final rather precipitous decline.

Le *et al.*'s results raise questions about authorial style and authorship attribution. It is a tenet of research on intrinsic methods of authorship attribution that each author has a unique and invariable "signature" that characterizes their writing — a signature that results from the combination of hundreds of largely unconscious predispo-

[5] Lancashire, 210–211.

[6] Conradi, 19, 654.

sitions and preferences that are reflected in every lexical and syntactic choice that the author makes in their writing.[7] Quantifiable elements of an author's signature include vocabulary size and distribution, average word and sentence length, distribution of prepositions and other function words, and frequency of various syntactic constructions.[8] The author's essential stylistic signature is taken to be largely independent of the topics upon which they write, and invariant, or largely so, even across different genres in which they write.

This tenet is challenged by the notion that diachronic changes in cognition, such as those of the early stages of Alzheimer's disease, might change these predispositions and preferences, or at least reduce the ability to act on them, and thus the author's signature might change to the extent that it is no longer recognizable as that of the author's earlier work. After all, some of the attributes commonly taken as components of an author's signature are attributes that change significantly in Alzheimer's disease — vocabulary richness and the frequency of certain syntactic constructions.[9] But on the other hand, it's also possible that many or most components of an

[7] Juola; Lancashire; Love.

[8] Grieve; Juola; Koppel, Schler, and Argamon; Stamatatos.

[9] Maxim and Bryan.

author's signature do not change, even in Alzheimer's disease,[10] and the essence of the author's individual style continues to shine through.

We can frame this as a question in authorship attribution or authorship verification: Using standard features for stylistic analysis, other than those known to change in Alzheimer's disease, can we recognize the work of an Alzheimer's-affected author as being written by the same author as that of his or her younger self, or do they appear to be two different authors? The use of standard features is important in this question. While we might find characteristics of the author's writing that are invariant even in cognitive decline, they do not answer our question about their stylistic signature unless they are characteristics that have been generally accepted in research on style as relevant to identifying individual authors.

It should be noted that this is not a question about authors' "late style" per se, which is a largely semantic or thematic notion related to the writer's outlook or view as it develops in their older years ("a new idiom" of "their work and thought" — Said, 6). Late style is not a matter of superficial linguistic realization, except only insofar as expression of the writer's new view might lead to some changes to preferences in surface realization.[11]

[10] We are speaking here, of course, of the earlier stages of Alzheimer's, in which the patient is still able to write, albeit perhaps poorly, as Murdoch and Christie did. In the later stages of the disease, when linguistic abilities may break down completely, the question becomes meaningless.

[11] Lancashire (218) points to Wyatt-Brown's characterization of the late style of Barbara Pym, which reflects both a new worldview and its surface realization: "Most reviewers and friends commented

In this paper, we use authorship attribution and verification methods to investigate whether an author's essential stylistic signature changes in Alzheimer's disease.

2. Authorship attribution and authorship verification

The hypothesis that we test is that intrinsic methods will distinguish "old" Agatha Christie and Iris Murdoch as different authors from their younger selves, but will not distinguish "old" P.D. James as different from her younger self. That is, in addition to the direct linguistic deficits that it induces, Alzheimer's disease changes the authors' style in detectable ways. The alternative (the null hypothesis) is that Alzheimer's disease does not change the writer's essential style, and that any differences seen are no larger than those seen over time in a healthily aging author. We test the hypothesis in two experimental frameworks: authorship attribution and authorship verification.

These two frameworks must be carefully distinguished.[12] In the framework of authorship *attribution*, we have a disputed text known to be written by one of a relatively small set of authors, and our task is to find the closest match of the stylistic signature of the text to those of the candidate authors: Was this mystery play writ-

upon the uncharacteristic undercurrents of depression and melancholy …. Furthermore, she herself described her desire to pare down her writing, to make things spare …" (Wyatt-Brown, 837). But it remains an open question as to whether Pym's new spareness or conciseness ("Pym had concentrated so passionately on making her point that later she had to add 12,000 words to make her manuscript publishable" (*ibid.*)) resulted in any marked change in the stylometrics of her writing.

[12] Koppel, Schler, and Bonchek-Dokow; Koppel, Schler, and Argamon.

ten by Shakespeare, Sheridan, or Shaw? We can answer the question by inducing a discriminative classifier from attested texts of the candidates and seeing which author it attributes the mystery text to. In the case of our particular hypothesis, this reduces to the simpler question of whether a classifier can be built at all to discriminate the works of the older author from those of her younger days. If the classifier cannot be built — that is, if no reliable and meaningful discriminating stylometric features can be found — then we may conclude that the author's style has not changed.

In the framework of authorship *verification*, the disputed text might or might not have been written by one particular author, but there are no specific known alternative candidates for authorship: Was this mystery play written by Shakespeare or not? In this situation, the notion of closest match does not apply, nor do we have any *a priori* notion of what would count as a close-enough match to the single candidate author; so the standard text-classification methods of intrinsic authorship attribution cannot be used. Instead, a method known as *unmasking* may be used. In the case of our hypothesis, we ask whether each work of the older author was or wasn't written by the "same" author as the works of her younger days.

Unmasking is a technique developed by Koppel and Schler.[13] It is based on the idea that if two texts were written by the same author, then any features a classifier finds that (spuriously) discriminate their authorship must be weak and few in number, and removal of these features from the classifier will seriously degrade its perfor-

[13] See also Koppel, Schler, and Bonchek-Dokow; Koppel, Schler, and Argamon.

mance (thereby “unmasking” the spurious discrimination). On the other hand, if the texts were written by different authors, then many features will support their (correct) discrimination and removal of even a few strongly discriminating features will not seriously harm the classifier. The method proceeds as follows: Classifiers are induced not only for the disputed text and its candidate author but also for the disputed text and other authors similar to the candidate. (Koppel and colleagues used a variety of nineteenth century English-language writers in their experiments.) The most discriminating features are progressively removed from each classifier and its performance is re-evaluated. If performance falls off far more rapidly for the candidate author, authorship is deemed to be verified. Recognizing when a “serious” decline in performance has occurred is a problem in meta-learning that we omit the details of here; the interested reader is referred to Koppel and Schler.

3. Experiments

3.1 Method

Data: We use the same data as Le *et al.*: 15 novels of Agatha Christie, including three of her very early novels (ages 28–34) and her final three novels from ages 80 to 82; 20 novels of Iris Murdoch, including seven of her first eight novels and her final two; and 15 novels of P.D. James, including her first eight and most recent five.

We divided each author’s novels into those of the author’s prime, those of the transition period when the author is approaching her late period, and those of her late period (see table 1). In the case of Murdoch and Christie, the late period contains novels in which the effects of Alzheimer's disease are possibly evident, as indicated by Le *et al.*'s results. The prime period contains novels in which there are presumed to be no effects of Alzheimer's. Novels of the transition period are those for which such judgements are less certain; in particular, as we noted earlier, Le *et al.* found possible indications of problems in Murdoch's novels even in her 50s, so we begin her transition period at age 50. For James, we made a division roughly similar to Christie's.

TABLE 1 HERE.

We divided each novel into large chunks and into small chunks. The large chunks were approximately 100 sentences long, breaking only at paragraph and chapter boundaries. For simplicity, we will refer to these as 100-sentence chunks, even though the number is approximate. We took these chunks as the basic unit of text in our preliminary study and in the authorship attribution study (sections 3.2 and 3.3 below). They averaged just over 4000 words — large enough to be representative, while giving us a sufficient number of textual units for each author for our experiments (see table 2). The small chunks averaged 542 words; they were made by breaking at the first paragraph boundary after 500 words. The small chunks were used in the authorship verification study (sections 3.2 and 3.4 below); for simplicity, we shall refer to them as 500-word chunks.

TABLE 2 HERE.

Classifier: We used support vectors machines (SVMs) as the classifiers in all our experiments. SVMs are classifiers commonly used in machine learning and data mining; they have shown competitive performance in authorship attribution research, and were used by Koppel *et al.* in their experiments on authorship verification. Specifically, we use the LibSVM classifier[14] of the Weka toolkit[15] in all our experiments.

Features: The 262 stylometric features that we used (table 3) are features that, broadly speaking, have been found to be effective in authorship attribution, including frequencies of function words, of characters and of character *n*-grams. Many are features that were tested for effectiveness in a study by Grieve, who found grapheme-based measures to be surprisingly effective. However, in this study we exclude stylometric features whose values Le *et al.* found to change in Alzheimer's disease — part-of-speech frequency and vocabulary richness — as it is the stability of the others that we wish to investigate. We do, however, include the frequency of the 20 most frequent part-of-speech bigrams in the text and the entropy of the part-of-speech tags (see table 3 for the definition) because, although these two features share a large amount of information with part-of-speech frequencies, they are effective in discriminating different authors[16] and Le *et al*. did not explore the correlation between longitudinal changes in these features and the presence of Alzheimer's disease.

[14] Chang and Lin.

[15] Hall *et al*.

[16] Hirst and Feiguina; Juola.

TABLE 3 HERE

In addition, for the preliminary study of authorship verification (section 3.2), we used, as a second feature set, the frequencies of the 250 most frequent words in the texts under consideration. We do this in order to follow Koppel *et al.*, who used the *n* words with the highest frequency in the data as a feature set in all their unmasking experiments, finding *n* = 250 to be a suitable value. We determined the set of words by the same method that they did, weighting frequency equally between the mystery text and the set of comparator texts. Many of these 250 most frequent words will be function words that are also used in the stylometric feature set (table 3).

3.2 Preliminary study: Authorship attribution and verification without regard to age

We verified in a preliminary study that our stylometric feature set is effective for authorship attribution and for authorship verification and that unmasking can discriminate between our three authors, without taking their age into account.

The first of these preliminary experiments used the authorship attribution framework. For each pair of authors, we took the set of all 100-sentence chunks of text by either author and trained an SVM classifier for the authors with the stylometric features. We also trained a three-way classifier with the texts of all three authors. The accuracy of classification, tested with 10-fold cross-validation, is shown in table 4; baseline is the accuracy obtained by always picking the author with the largest number of texts in the data.

TABLE 4 HERE

We see that these stylometric features are effective for authorship attribution on this set of writers. The accuracy is somewhat lower than is sometimes seen in authorship attribution studies, but this may be attributed to our deliberate avoidance of vocabulary and part-of-speech features.

The second of these preliminary experiments used the authorship verification framework. We used both our stylometric feature set (table 3) and for comparison, following Koppel *et al.*, a feature set of the 250 most frequent words in the texts, as described in section 3.1 above. And also following Koppel *et al.*, we used our smaller 500-word chunks of text in this experiment.

For each of the 50 novels and for each author, we used unmasking to test whether that novel was written by that author. Thus there were 150 test cases. The results are shown in table 5. Both feature sets achieved an accuracy of approximately 89% (specifically, the 250-word feature set was correct in one case more than the stylometric feature set), well above the baselines of 40% (always saying that the book was written by the most-frequent author, Iris Murdoch), and 50% (making the decision of same or different author at random with equal probability) . Thus, we see that unmasking is an effective authorship verification method for these data, and that our standard stylometric feature set is as effective in unmasking as Koppel *et al.*'s and is therefore suitable for use in our age-based unmasking experiments below.

TABLE 5 HERE

3.3 Discriminating ages in the authorship attribution framework

We next tested whether we can distinguish the two Alzheimer's-affected authors from their younger selves, with P.D. James as a control subject. We did this by seeing whether it was possible to build for each author a binary classifier based on stylometric features that could discriminate her prime-period work from her late-period work. (We ignored data from the transition period in order to separate the author's work into two clear cases.) If our hypothesis is correct, we should be able to do this for Murdoch and Christie but not James. As in our preliminary study, we used 100-sentence chunks of text for each author; the number of chunks for each period is shown in table 2.

The results are shown in table 6. The baseline for comparison is the accuracy obtained by always choosing the more-frequent period, which is the prime period in all cases. Because Murdoch and Christie have much more work in this period, this is a high baseline for them. *P* is the probability that the classifier is operating at the baseline chance level, as determined by the exact binomial test (using the R statistical software package[17]). We see that the difference is significant for Christie and James while barely even approaching mild significance for Murdoch.

TABLE 6 HERE

---

[17] R Development Core Team.

Thus, in accordance with our hypothesis, we are able to use stylometric features to discriminate Christie's work by age in the authorship attribution framework, but, contrary to our hypothesis we could also do so for James and we could not do so for Murdoch.

3.4 Discriminating ages in the authorship verification framework

We used Koppel *et al.*'s unmasking technique to determine whether each author's late-period work could be verified as having the same author as her prime-period work. If our hypothesis is correct, we should be able to verify this for James but not for Murdoch or Christie.

Separately for each author, we did the following with our stylometric features and the 500-word-chunk dataset: For each novel by that author, except for those in the transition period, and for each period, prime and late, we used unmasking to test whether that novel was written in that period. There are two baselines in this experiment. As in the classification experiment (section 3.3, table 6), an informed baseline is verifying authorship as the more-frequent "author", the prime period, in all cases — that is, asserting identity of authorship in all cases when the prime period is the target and denying it in all cases when the late period is the target. But such an informed baseline is perhaps unrealistic in the authorship verification framework, so we also consider the uninformed baselines of equiprobable random choice and of always denying identity of authorship (which would be a high baseline in the realistic situation in which most work tested is not that of the target author). For our data, these baselines are equivalent, both giving 50% accuracy.

The results are shown in Table 7. *P* is the probability that the classifier is operating at the baseline chance level, as determined by the exact binomial test. We see that in all three cases, unmasking is operating at chance level compared to the informed baseline; in fact, for Murdoch and Christie, it performs markedly *worse* than just choosing the prime period for all cases. That is, unmasking cannot find any features in the stylometric feature set that reliably discriminate the age of any of the authors. This contrasts with our results for binary SVM classification, where, using the same feature set, we were able to make a significant discrimination for two of our three authors. On the other hand, unmasking did achieve a significantly better result than the uninformed baseline for Murdoch, and approached significance for James; it did not do so for Christie. This result is contrary to that of SVM classification (table 6), where Murdoch was the most poorly discriminated author.

TABLE 7 HERE

Thus, contrary to our hypothesis, we were not able to use stylometric features to discriminate Christie's and Murdoch's work by age in the authorship verification framework, except that for Murdoch we could discriminate better than an uninformed baseline. However, in accordance with our hypothesis, we could not discriminate James either, though we came very close to significance against the uninformed baseline.

4. Discussion

The inconsistency in our results draws attention to methodological issues. On one hand, binary SVM classification — a well-established method — was able to make discriminations between the prime and late periods of Christie and James using well-established stylometric features, and came close to doing so for Murdoch. From this we can conclude that these authors' styles changed in old age, but because this was so for P.D. James as well as our Alzheimer's-affected authors, we can draw no conclusions about changes that might be directly due to Alzheimer's disease. It is notable that we found changes in James's writing with regard to stylometric features even though Le *et al.* found no significant changes in her writing with regard to features that are known to commonly change in Alzheimer's disease. This supports the idea that these sets of features do not co-vary and hence, in principle at least, the stylometric features could largely remain unchanged even in Alzheimer's, even if they didn't do so in the present cases of Christie and, to a degree, Murdoch.

On the other hand, unmasking could not discriminate the authors by age very well, or could not do so at all, depending on what baseline was chosen. Unmasking is a less well established method; although Koppel *et al.* report good results with it, and it worked well in our preliminary study, it has not been widely used by other researchers. Possibly, in view of the SVM results, it is not suitable for this kind of authorship verification; that is, the features do not behave in the manner that the method assumes.

Thus our hypothesis that Alzheimer's disease would lead to changes in an author's essential style, distinct from the linguistic changes seen in cognitive decline, is neither supported nor falsified. Rather, one of our underlying axioms — the invariance of an author's essential stylistic signature, as reflected in standard stylometric features, in the absence of cognitive decline — is brought into question. Most research on authorship attribution by stylometric features focuses on the author's work as a whole, seeking features that reflect differentiation from other authors and hence similarity among the author's various works rather than differences. Determining the effects of cognitive decline on an author's essential style first requires a better understanding of the norms of diachronic changes in style.[18]

[18] See Lancashire for an example of research aimed at developing this understanding.

Acknowledgements

This work was supported financially by the Natural Sciences and Engineering Research Council of Canada. We are grateful to Michelle Tran for conducting a pilot study of unmasking on these data.

References

Chang, Chih-Chung and Chih-Jen Lin. "LIBSVM: A library for support vector machines." *ACM Transactions on Intelligent Systems and Technology*, 2(3) (2011): article 27.

Conradi, Peter J. *Iris Murdoch: A Life*. New York: Norton, 2001.

Garrard, Peter, Lisa M. Maloney, John R. Hodges, and Karalyn Patterson. "The effects of very early Alzheimer's disease on the characteristics of writing by a renowned author." *Brain* 128(2) (2005): 250–260.

Grieve, Jack. "Quantitative authorship attribution: An evaluation of techniques." *Literary and Linguistic Computing* 22(3) (2007): 251–270.

Hall, Mark, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten. "The WEKA data mining software: An update." *SIGKDD Exploration*s 11(1) (2009): 10–18.

Hirst, Graeme and Ol'ga Feiguina. "Bigrams of syntactic labels for authorship discrimination of short texts." *Literary and Linguistic Computing* 22(4) (2007): 405–417.

Juola, Patrick. "Authorship attribution." *Foundations and Trends in Information Retrieval* 1(3) (2006): 233–334.

Koppel, Moshe and Jonathan Schler. "Authorship verification as a one-class classification problem." *Proceedings of the 21st International Conference on Machine Learning*, Banff, 489–495.

Koppel, Moshe, Jonathan Schler, and Elisheva Bonchek-Dokow. "Measuring differentiability: unmasking pseudonymous authors." *Journal of Machine Learning Research* 8 (2007): 1261–1276.

Koppel, Moshe, Jonathan Schler, and Shlomo Argamon. "Computational methods in authorship attribution." *Journal of the American Society for Information Science and Technology* 60(1) (2009): 9–26.

Lancashire, Ian. *Forgetful Muses: Reading the Author in the Text.* Toronto: University of Toronto Press, 2010.

Le, Xuan, Ian Lancashire, Graeme Hirst, and Regina Jokel. "Longitudinal detection of dementia through lexical and syntactic changes in writing: A case study of three British novelists." *Literary and Linguistic Computing* 26(4) (2011): 435–461.

Love, Harold. *Attributing Authorship: An Introduction.* Cambridge University Press, 2002.

Maxim, Jane and Karen Bryan. *Language of the Elderly: A Clinical Perspective.* London: Whurr, 1994.

Nicholas, Marjorie, Lorraine K. Obler, Martin L. Albert, and Nancy Helm-Estabrooks. "Empty speech in Alzheimer's disease and fluent aphasia." *Journal of Speech and Hearing Research* 28 (1985): 405–410.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2010.

Said, Edward W. *On Late Style: Music and Literature Against the Grain.* New York: Pantheon Books, 2006.

Stamatatos, Efstathios. "A survey of modern authorship attribution methods." *Journal of the American Society for Information Science and Technology* 60(3) (2009): 538–556.

Wyatt-Brown, Anne M. "Late style in the novels of Barbara Pym and Penelope Mortimer." *The Gerontologist* 28(6) (1988): 835–839.

Table 1. Novels studied, with author's age at time of writing and categorization into periods: prime, transition, or late.

Iris Murdoch

Prime-period novels (9)

35	*Under the Net*
36	*The Flight from the Enchanter*
39	*The Bell*
42	*A Severed Head*
43	*An Unofficial Rose*
44	*The Unicorn*
45	*The Italian Girl*
47	*The Time of the Angels*
49	*The Nice and the Good*

Transition-period novels (9)

50	*Bruno's Dream*
51	*A Fairly Honorable Defeat*
54	*The Black Prince*
55	*The Sacred and Profane Love Machine*
57	*Henry and Cato*
59	*The Sea, the Sea*
64	*The Philosopher's Pupil*

66 *The Good Apprentice*

68 *The Book and the Brotherhood*

Late-period novels (2)

74 *The Green Knight*

76 *Jackson's Dilemma*

Agatha Christie

Early-period novels (10)

28 *The Mysterious Affair at Styles*

32 *The Secret Adversary*

34 *The Murder of Roger Ackroyd*

43 *Murder on the Orient Express*

47 *Appointment with Death*

50 *Curtain*

51 *Towards Zero*

59 *A Murder is Announced*

63 *Destination Unknown*

67 *Ordeal by Innocence*

Transition-period novels (2)

72 *The Clocks*

76 *Endless Night*

Late-period novels (3)

80 *Nemesis*

81 *Elephants Can Remember*

82 *Postern of Fate*

P.D. James

Prime-period novels (7)

42 *Cover Her Face*

43 *A Mind to Murder*

47 *Unnatural Causes*

51 *Shroud for a Nightingale*

52 *An Unsuitable Job for a Woman*

55 *The Black Tower*

57 *Death of an Expert Witness*

Transition-period novels (4)

60 *Innocent Blood*

66 *Taste for Death*

72 *The Children of Men*

77 *A Certain Justice*

Late-period novels (4)

81 *Death in Holy Orders*

83 *The Murder Room*

85 *The Lighthouse*

88 *The Private Patient*

Table 2. Size of data for each author: Number of 100-sentence text chunks in each period, and average chunk size in words.

	Period				**Avg size**
	Prime	**Transition**	**Late**	**Total**	**(words)**
Iris Murdoch	219	43	49	311	4427
Agatha Christie	244	94	72	410	3033
PD. James	126	23	100	249	5161
All				949	4051

Table 3. Stylometric features used. The numbers in parentheses denote the number of features in the category; categories not so marked are a single feature.

1. *Lexical features* (10)
    a. Frequencies of function words
        i. Prepositions
        ii. Pronouns
        iii. Determiners
        iv. Conjunctions
        v. Modal auxiliaries
        vi. Primary verbs: *be, have, do*
        vii. Adverbs
        viii. *To*
    b. Frequencies of hapax legomena
    c. Frequencies of hapax dislegomena

2. *Character features* (231)
    a. Count of all alphabetic characters
    b. Count of all digit characters
    c. Upper-/lower-case character count (2)
    d. Letter frequencies (26)
    e. Count of punctuation marks

f. Frequencies of the most frequent letter *n*-grams (100 bigrams and 100 trigrams)

3. *Syntactic features* (21)

a. Frequencies of the most frequent part-of-speech bigrams (20)

b. The entropy H of part-of-speech tags, given by $H = -\sum_{i=1}^{N} p_i \log_2 p_i$, where p_i is the probability of the ith part-of-speech tag and N is the number of different tags.

Table 4. Accuracy (%) of classification of each pair of authors studied, and of three-way classification of all authors, using novels of all the authors' periods, with the stylometric features of table 3 and an SVM classifier. The baseline is choosing the more-frequent or most-frequent category.

Authors	**Accuracy**	**Baseline**
Murdoch–Christie	90.3	55.6
Murdoch–James	98.0	55.5
Christie–James	93.6	61.0
Murdoch–Christie–James	87.9	41.0

Table 5. Accuracy (%) of unmasking technique for author verification on each novel of the authors studied (using novels of all the authors' periods), with the stylometric features of table 3 and with the 250-word feature set, using an SVM classifier. The baselines are 40% accuracy (choosing the most-frequent author) and 50% accuracy (equiprobable random choice between *same author* and *different author*). *P* is the probability that the results of the classifier are not significantly better than the higher baseline.

Feature set	**Number of cases**	**Correct**	**Incorrect**	**Accuracy**	***P***
Stylometric	150	133	17	88.7	$\ll .001$
250-words	150	134	16	89.3	$\ll .001$

Table 6. Accuracy (%) of binary SVM classification of 100-sentence chunks of each author's work into prime period and late period. Baseline is choosing the more-frequent category, i.e., the prime period in each case. *P* is the probability that the results of the classifier are not significantly better than the baseline.

	Number of cases	**Correct**	**Incorrect**	**Accuracy**	**Baseline**	***P***
Murdoch	268	227	41	84.7	81.7	.115
Christie	316	277	39	87.7	77.5	$\ll$.001
James	226	170	56	75.2	55.8	$\ll$.001

Table 7. Accuracy (%) of unmasking technique for verification of each author's work as prime period or late period. Baseline 1 is the informed baseline of choosing the more-frequent category, i.e., the prime period in each case; baseline 2 is uninformed equiprobable random choice. For each, *P* is the probability that the results of the classifier are not significantly better than the baseline.

	No of cases	**Correct**	**Incorrect**	**Accuracy**	**Baseline 1**	***P***	**Baseline 2**	***P***
Murdoch	22	16	6	72.7	81.7	.908	50.0	.026
Christie	26	16	10	61.5	77.5	.981	50.0	.164
James	22	15	7	68.2	55.8	.170	50.0	.067