

THE AUTHORSHIP OF THE *HISTORIA AUGUSTA*:
TWO COMPUTER STUDIES *

By IAN MARRIOTT

I

The most widespread and best-known use of computers in stylistic analysis has been in the attempts to resolve questions of disputed authorship. Here, the computer's ability to store the whole known corpus of an author's work and to compare it with the disputed work in many ways which would be inexpressibly tedious, if not impossibly lengthy, by hand, is an invaluable aid.¹ Many studies have been carried out on classical texts where the authorship is doubtful,² but these studies have most often had undisputed texts against which a disputed work could be compared. The *Historia Augusta* presents a different situation: there are six 'authors' to whom varying amounts of the work are assigned. The following table lists the 'authors' and the lives which are attributed to each of the six.

'Aelius Spartianus'	'Iulius Capitolinus'	'Vulcacius Gallicanus'	'Aelius Lampridius'	'Trebellius Pollio'	'Flavius Vopiscus'
Hadrianus	Antoninus Pius	Avidius Cassius	Commodus	Valeriani Duo	Aurelianus
Aelius	Marcus Aurelius		Diadumenian	Gallieni	Tacitus
Didius Iulianus	Verus		Elagabalus	Tyranni Triginta	Probus
Septimius Severus	Pertinax		Alexander Severus	Claudius	Quadrigae Tyrannorum
Pescennius Niger	Clodius Albinus				Carinus
Caracalla	Opilius Macrinus				
Geta	Maximini Duo				
	Gordiani Tres				
	Maximus and Balbinus				

None of the six is attested elsewhere, and, therefore, the emphasis of any study must be on assessing whether or not the differences and similarities between the six 'authors' are indicative of plural or singular authorship of the work as a whole. In this, comparison

* The research on which this paper is based was carried out at University College London under the supervision of Mr. R. I. Ireland.

¹ See in particular A. Ellegard, 'A Statistical Method for Determining Authorship: The Junius Letters, 1769-1772', *Gothenburg Studies in English* 13 (1962); H. Coppens-Ide, 'Authorship Problems and The Computer', *L.A.S.L.A. Revue* 3 (1971), 187; F. Mosteller and D. L. Wallace, *Inference and Disputed Authorship: The Federalist* (1964); G. U.

Yule, *The Statistical Study of Literary Vocabulary* (1944).

² e.g. L. Brandwood, 'Analysing Plato's Style with an Electronic Computer', *BICS* 3 (1956), 45; A. Q. Morton, 'The Authorship of Greek Prose', *Journ. Roy. Stat. Soc.* 128A (1965), 169; *St Paul, The Man and The Myth* (1966); *It's Greek to the Computer* (1971); S. V. F. Waite, 'Approaches to the Analysis of Latin Prose, Applied to Cato, Livy and Sallust', *L.A.S.L.A. Revue* 2 (1970), 91.

with other works is vital since these will act as controls. Some work on linguistic and stylistic resemblances between the putative six authors has been carried out,³ but what follows represents, so far as is known, the first systematic application of computer techniques to this much discussed problem.

It should be noted that, although the findings presented here may have relevance for the dating of the *Historia Augusta*, this study is solely concerned with the problem of authorship.

II

The first study concerns the distributions of sentence lengths exhibited by the 'authors' of the *Historia Augusta*.

For the purposes of the analysis, a sentence was defined as a sequence of words terminated by a full-stop, colon or interrogation mark. The colon was included because W. C. Wake has demonstrated that, if a sentence is defined as a group of words ended by a full-stop or interrogation mark, then, although there was a good deal of agreement between the editors of texts, there could be a significant difference between them; if the colon is included, the differences between editors cease to have any such significance.⁴

The statistical measures applied to the results are the *arithmetic mean* (defined as

$$\bar{x} = \frac{\sum x}{N}$$

which indicates that all the values of the distribution are to be summed and then divided by the number of items: in this case, this is simply the total number of words divided by the total number of sentences) and the *standard deviation* (defined as

$$s = \sqrt{\frac{\sum(x - \bar{x})^2}{N}}$$

which shows how closely the individual values of a distribution group around the mean).

The most effective way to approach the problem of the *Historia Augusta* is to consider first the six 'authors' and then the individual lives. Table 1 presents the statistics for the six authors (plus the overall figures for the entire text) and for several control works. The control works used here are the anonymous *De Rebus Bellicis*, a large selection from the *Codex Theodosianus* and six books, chosen at random, from Ammianus Marcellinus. These figures are also presented graphically in Fig. 1.

Before any assessment can be made of these statistics, two important topics must be considered. The first of these concerns some of the fundamental theory behind quantitative stylistics, while the second concerns statistical significance.

Firstly, then, style is a flexible entity, one best described in terms of tendencies. It is a *probabilistic concept*: in an ideal world, the occurrence of one phenomenon is not unequivocally predetermined by the existence of a given condition. The phenomenon only has a *probability* of occurrence and in an ideal ensemble, the expectancy of any of the series of phenomena A, B and C occurring could be expressed by a *probability distribution*. Empirically, however, the phenomena A, B and C can also occur in non-ideal worlds, where the values of the probability distribution will be liable to random fluctuations and will not be stable. Thus, the occurrences of the phenomena A, B and C have to be described in terms of *frequency distributions*, which will fluctuate in a certain, admissible interval around the probability distribution.

Thus, a frequency distribution describes the physical manifestations of an author's overall style, his preferences for one or another mode of expression. Variations will be expected, and this leads immediately to the question of *statistical significance*: how much variation can be attributed to random factors, and how much is due to real discrepancies

³ Note especially P. White, 'The Authorship of the *Historia Augusta*', *JRS* 57 (1967), 115, esp. 125-8; J. N. Adams, 'On the Authorship of the *Historia Augusta*', *CQ* 22 (1972), 186.

⁴ 'Sentence Length Distributions of Greek Authors', *Journ. Roy. Stat. Soc.* 120A (1957), 331. Cf. T. Janson, 'Word, Syllable and Letter in Latin', *Eranos* 65 (1967), 49.

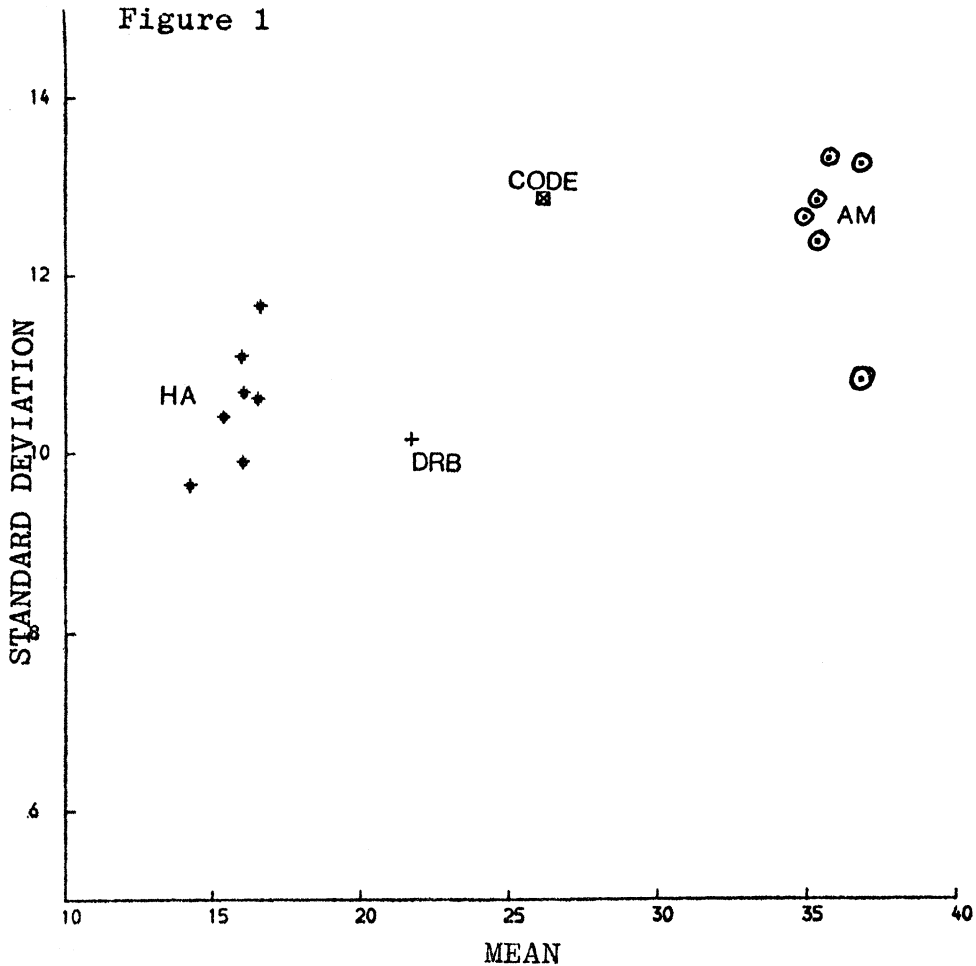


FIG. 1. SENTENCE-LENGTH DISTRIBUTIONS FOR THE SIX *HISTORIA AUGUSTA* 'AUTHORS' AND FOR THE CONTROL TEXTS: HA INDICATES THE SIX *HISTORIA AUGUSTA* 'AUTHORS', DRB THE *DE REBUS BELLICIS*, CODE THE *CODEX THEodosIANUS* AND AM AND DOT ONE BOOK OF AMMIANUS MARCELLINUS. *Copyright reserved*

between texts or authors? The assessment of variation is made by using *significance tests*, which express in probabilities whether differences may, or may not, be statistically significant.

One of the most widely-used significance tests is the *standard error of difference*. This is defined as

$$\text{S.E.D. } (\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

and means that if the difference between the means is greater than two standard errors, it is probably significant, that is, not very likely to have arisen by chance; if the difference between the means exceeds three standard errors, it is definitely significant, the associated probability being of the order of less than one-half of 1 per cent that so great a difference could arise by chance.⁵

If this test is applied, the picture presented in Fig. 1 is confirmed: internally, the *Historia Augusta* is homogeneous. The six 'authors' show no significant difference

⁵ The calculation is as follows: subtract the mean of one text from the mean of another; then, for each of these texts, take the square of the standard deviation and divide this by the number of items,

and add these two values together; finally, take the square of this sum to give the standard error of difference between the means.

among themselves, while they are totally distinct from any of the control texts. The method receives validation from the fact that all six books of Ammianus are also statistically homogeneous; the *Codex* and the *De Rebus Bellicis* stand by themselves and there is no overlap between any of the texts.

Bearing this in mind, what can be said about the individual lives of the *Historia Augusta*? Table 2 presents the figures for the lives, and these are also presented graphically in Fig. 2 (1 means a life assigned to 'Spartianus', 2 to 'Capitolinus', 3 to 'Gallicanus', 4 to 'Lampridius', 5 to 'Pollio' and 6 to 'Vopiscus').

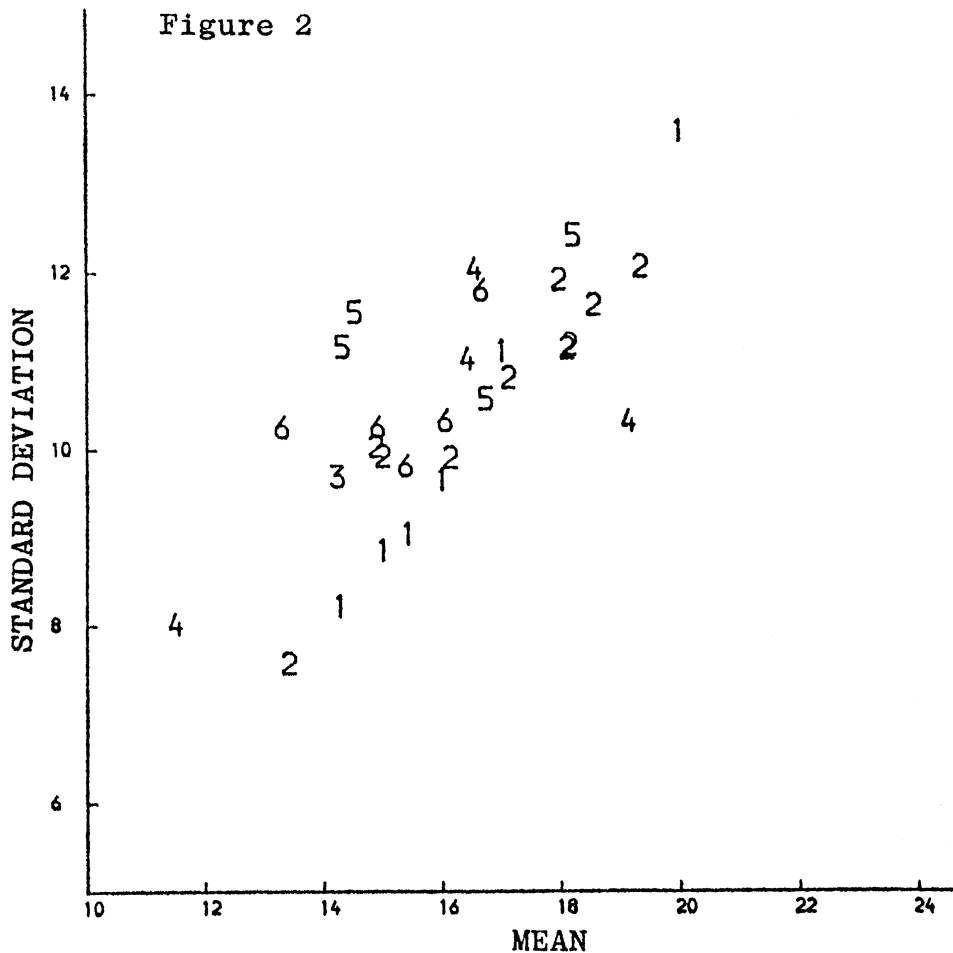


FIG. 2 SENTENCE-LENGTH DISTRIBUTIONS FOR THE INDIVIDUAL LIVES OF THE *HISTORIA AUGUSTA*: 1 INDICATES A LIFE ASSIGNED TO 'AELIUS SPARTIANUS', 2 TO 'IULIUS CAPITOLINUS', 3 TO 'VULCACIUS GALLICANUS', 4 TO 'AELIUS LAMPRIDIUS', 5 TO 'TREBELLIIUS POLLIO' AND 6 TO 'FLAVIUS VOPISCUS'. *Copyright reserved*

It is immediately noticeable that the values for mean and standard deviation cover a wide range. However, using the same significance test as before, the lives are almost overwhelmingly homogeneous. Indeed, when compared against their respective 'authors', there are no significant differences to be found, and when compared against the control works, the lives are totally distinct. Yet, there are differences between some of the lives and to leave matters thus would be simplistic—the question of why these differences occur must be answered. At this point, it becomes apparent that the answer to this question, and, perhaps, to the whole problem of the *Historia Augusta*, lies not in the number of authors, but in the different levels and types of composition within the work.

Although the six 'authors' claim to have consulted an enormous number of sources (some thirty-five of which are elsewhere unattested), it is generally agreed that, in the early lives, one major source can be detected.⁶ This enables a division to be made: the lives from Hadrian to Geta represent one section, those from Macrinus another. The first section can be further subdivided into two groups.

The first group, known as the 'primary' lives, comprises the biographies of Hadrian, Antoninus Pius, Marcus Aurelius, Verus, Commodus, Pertinax, Didius Iulianus, Septimius Severus and Caracalla. This leaves a group of what are known as 'secondary' lives, which cover Aelius, Avidius Cassius, Pescennius Niger, Clodius Albinus and Geta. The major source for the lives of the emperors appears not to have contained separate sections for the minor figures (the 'secondary' lives concern one Caesar, one prince and three pretenders), the details for whom are either drawn from the 'primary' lives to which they are appended, or are invented, or both. The lives from Macrinus to the end are largely the result of invention, this process becoming more noticeable as the biographies progress, and being indicated, as it is in the 'secondary' lives, by a growing proliferation of fictitious documents, speeches, decrees, etc.

If this is taken into account, the swarm of points in Fig. 2 take on a pattern. To clarify this pattern, Fig. 3 presents the same information, but, here, the numbering is different: 1 indicates a 'primary' life, 2 a 'secondary' and 3 a life from the later group.

With the exception of the *Verus*, the 'primary' lives group together in the bottom left-hand corner of the graph. A closer inspection of these points to a further subdivision within them into three groups, each classifiable by its proximity to the centre of the graph.

The first group, that furthest from the centre, contains the lives of Didius Iulianus, Pertinax and Commodus, the second those of Septimius Severus and Caracalla, the third those of Hadrian, Antoninus Pius and Marcus Aurelius.

Although statistically homogeneous, the variation in the 'primary' lives is evident, and if this is to be accounted for, must be attributed either to variation in the common source or to how the 'authors' handled the source. Since the former possibility is impossible to investigate, what can be said about the latter?

There is a tension between the 'primary' and 'secondary' lives by which elements characteristic of the latter find their way into the former, and which manifests itself to varying degrees. The extent to which this process affects the 'primary' lives seems to explain the variation within them.

The 'primary' lives closest to the centre of the graph in Fig. 3 are those of Hadrian, Antoninus Pius and Marcus Aurelius. These show signs of much manipulation, particularly in the form of appendages and additions. In the *Hadrian*, for example, chs. 14, 5 to 16, 7 form an appendage on the emperor's personal life, a device also evident in the *Antoninus Pius* in chs. 7, 5-12 and 10 to 11, and in the *Marcus Aurelius* in chs. 28 and 29. Other characteristics of the 'secondary' and later lives are: the list of largely bogus names in *Antoninus Pius*, 12, and *Marcus Aurelius*, 2 and 3; the oracles at *Hadrian*, 2, and the omens at *Antoninus Pius*, 3 and 9; the scandals described in *Marcus Aurelius*, 19. Chs. 16, 3 to 18 of the *Marcus Aurelius* are held to derive, not from the main source, but from Eutropius VIII, 11-XIV, 2.

The same process can be seen in the lives of Septimius Severus and Caracalla, but to a lesser degree. In the *Severus*, this can be seen in ch. 13 or in the invocation to Diocletian in ch. 20, in which place there is the only instance of the citation of a bogus authority in the 'primary' lives. Chs. 8 to 11 of the *Caracalla* form an appendage and, in the *Severus* again, chs. 17, 5 to 19, 4 are thought to derive from Aurelius Victor, 20, 1 to 31.

The lives furthest from the centre of the graph, the *Didius Iulianus*, *Pertinax* and *Commodus*, show little evidence of any such tampering, and it is therefore likely that they best represent the manner of the common source.

This general picture is disturbed by the position of the *Verus*, which is located on the top right-hand side of the graph, much removed from the other 'primary' lives.

Although Verus was indeed Augustus for the years A.D. 161-9 (and as such would be

⁶ For a detailed discussion of this question see T. D. Barnes, *The Sources of the Historia Augusta*, Coll. Latomus CLV (1978).

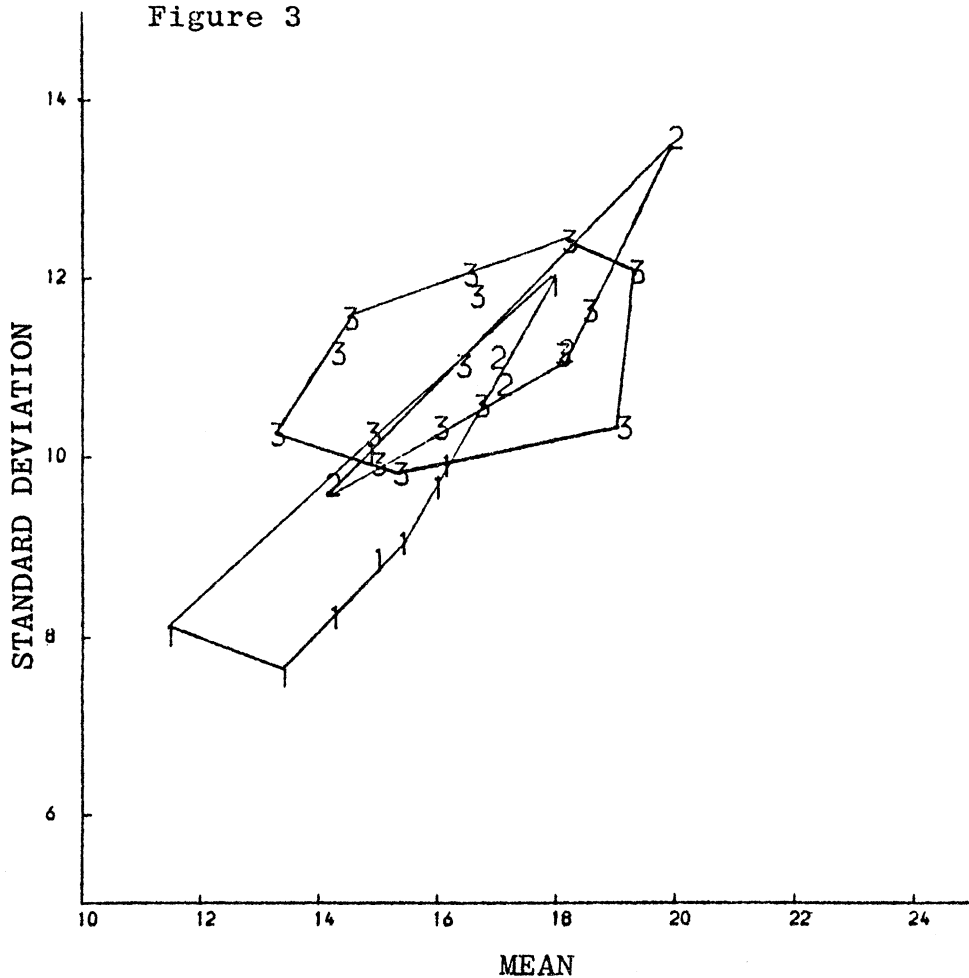


FIG. 3. SENTENCE-LENGTH DISTRIBUTIONS FOR THE INDIVIDUAL LIVES OF THE *HISTORIA AUGUSTA*, BUT SHOWING THE DIVISION INTO PRIMARY, SECONDARY AND LATER LIVES, WITH *VERUS* BEING CLASSIFIED AS PRIMARY: 1 INDICATES A PRIMARY, 2 A SECONDARY AND 3 A LATER LIFE. THE OUTLYING MEMBERS OF EACH GROUP ARE CONNECTED BY LINES TO SHOW THE CLOSENESS OF THE GROUPING. *Copyright reserved*

expected to be represented in the common source), nevertheless, scrutiny of this life reveals that, structurally, it is considerably like a 'secondary' life: it has a preface and contains much reworked material from the *Marcus Aurelius*; six out of the eleven chapters are given over to scandal and gossip (chs. 4 to 9).⁷ This raises the possibility that, despite Verus' status as Augustus, he may not have received separate treatment in the common source. If this life is considered as 'secondary', the configuration as shown in Fig. 4 results: the pattern is now definite and relates to the results very well.

Now, as the 'primary' lives cluster in the bottom left-hand corner of the graph, so the 'secondary' lives tend towards the upper right-hand corner. Unlike the 'primary' lives, their positions are much more varied and there is considerable overlap with the later lives. This is as expected, since the manner of composition in the 'secondary' lives corresponds to a great degree with that used in those from the *Macrinus* to the end.

The later, largely fictitious lives vary widely across the graph, but there is one remarkable point. If singular authorship is accepted, 'Vopiscus' represents the culmination of the work and so may be reasonably expected to represent the personality behind it best. The lives assigned to him occupy central positions on the graph and are the most homogeneous of all.

⁷ See T. D. Barnes, 'Hadrian and Lucius Verus', *JRS* 57 (1967), 65, on pp. 66-74.

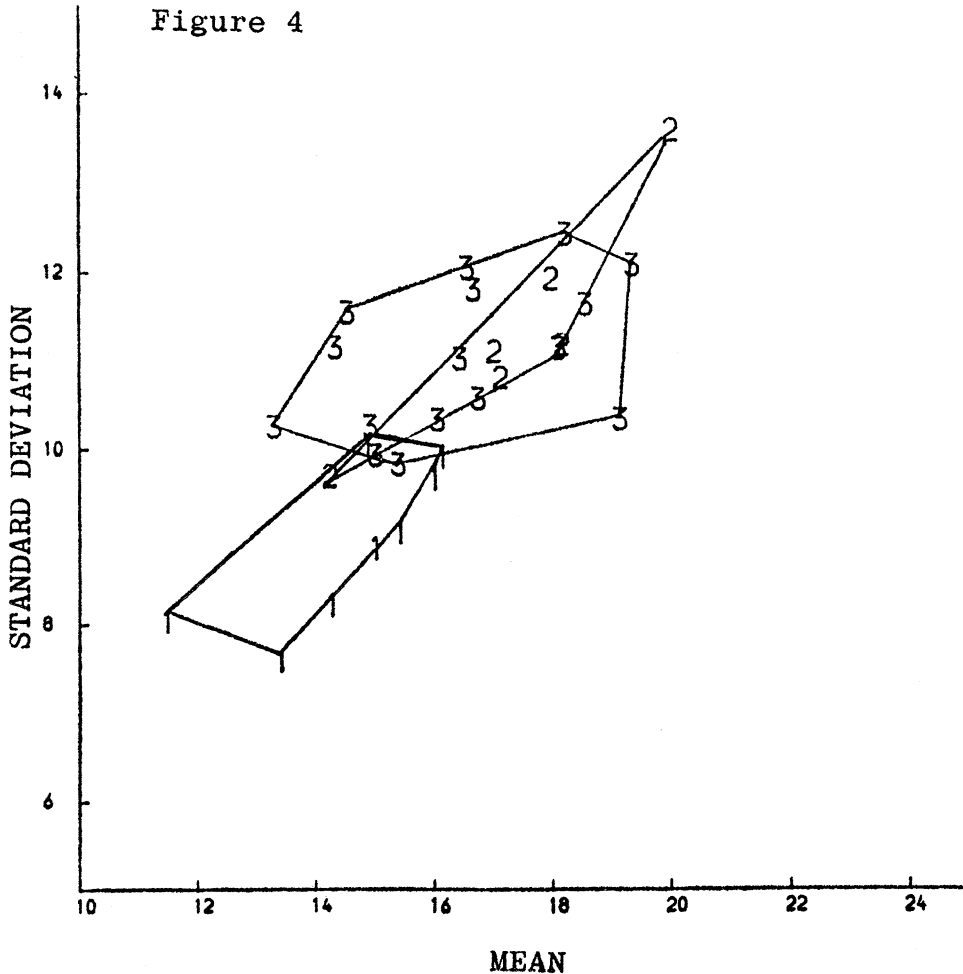


FIG. 4. SENTENCE-LENGTH DISTRIBUTIONS FOR THE INDIVIDUAL LIVES OF THE *HISTORIA AUGUSTA*, AGAIN SHOWING THE DIVISION INTO PRIMARY, SECONDARY AND LATER LIVES, BUT HERE CLASSIFYING *VERUS* AS A SECONDARY LIFE: 1 INDICATES A PRIMARY, 2 A SECONDARY AND 3 A LATER LIFE. THE OUTLYING MEMBERS OF EACH GROUP ARE CONNECTED BY LINES TO SHOW THE CLOSENESS OF THE GROUPING. *Copyright reserved*

III

The second study carried out also concentrates on the sentence: it examines the grammatical types of word used in initial and final positions in a sentence—these types were defined as noun, pronoun, verb, adverb, adjective, participle, preposition, numeral and enclitic.

Table 3 contains the results of this test, with the figures being the totals for each word-class expressed as a percentage of the number of sentences in the work.

In this study, no semantic associations were made with the words, because, whereas the meaning of a word may be more indicative of subject-matter, its type, it is felt, will rather reflect the writer's stylistic habits and choices.

A writer can select points within a sentence on which to lay particular emphasis, but the two positions which lend themselves naturally to emphasis are the beginning and the end. When word-order is free, the types of words used in these positions will give a view of a writer's preferences, and he should be consistent in the relative proportions of the various types which he uses.

The figures of Table 3 for the six 'authors' and the control texts are presented in Fig. 5, where these shapes are what are sometimes called *ray diagrams*.

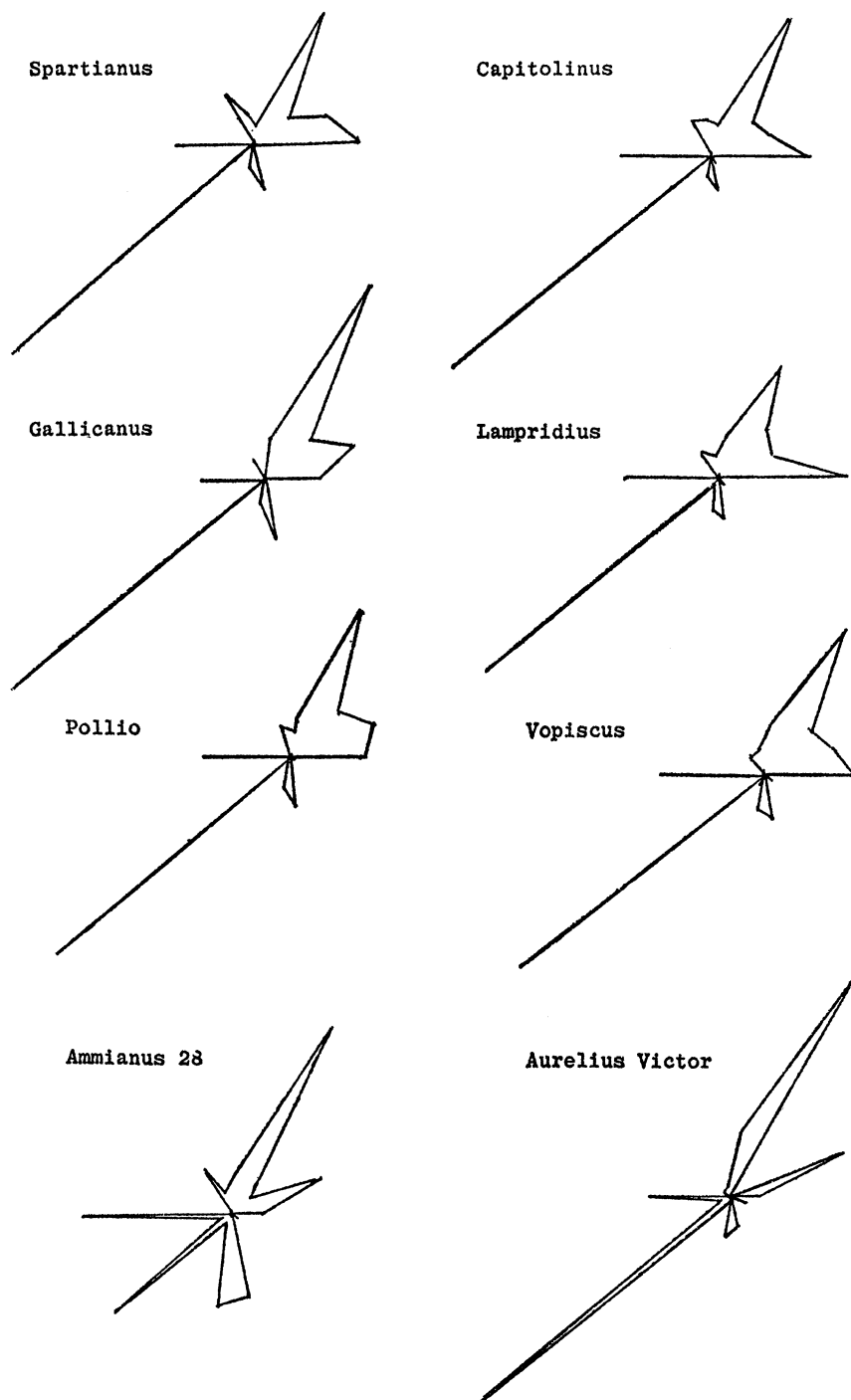


FIG. 5. RAY DIAGRAMS FOR THE SIX *HISTORIA AUGUSTA* AUTHORS AND FOR THE CONTROL TEXTS (AMMIANUS BK. 28 AND AVRELIUS VICTOR, *DE CAESARIBUS* Copyright reserved)

Since there are eighteen word-classes in all, nine for initial and nine for final words, a circle with eighteen equiangular radii is drawn. The frequency of occurrence, expressed as a percentage, is measured along the appropriate radius (working in an anticlockwise direction) and the point is marked. Then, the points are joined. Thus a direct visual

comparison can be made between the diagrams, since, in a very real way, a shape is the author or text it portrays.

Consideration of the diagrams reveals a remarkable situation: not only are the shapes for the *Historia Augusta* strikingly similar—coinciding almost exactly on more than one value—but, even allowing for considerable variation of styles, they are very different from either of the controls.

A closer examination of the results is rewarding.

Firstly, Aurelius Victor, in the text analysed, never opens a sentence with a verb: his favourite initial word-type is the adverb—over 50 per cent of his sentences open with one. Excellent examples of this habit are seen in chs. 11, 14, 15, 28 and 33.

Connected to this use of adverbs is Victor's tendency to attach '-que' to them. Indeed, he has a great liking for this enclitic, as is shown by the fact that nearly 3 per cent of his sentences end in it and the fact that only for him does the total for '-que' outweigh the more usual high-frequency connective 'et' by a ratio of almost five to one, whereas in all the other texts this ratio is reversed.

Of the other possible word-types, only a restricted few are used heavily: Victor favours, in initial position, pronouns as well as adverbs, and, in final position, verbs.

The distinctive shape for Victor is created by these choices.

Ammianus Marcellinus, however, presents a very different picture. Overall, the shape is more balanced, with only the score of 42 per cent for initial adverbs disturbing the diagram. The use of initial and final word-types is much more evenly spread: the score for final verbs is half that of any of the other writers and he exhibits by far the highest totals for nouns, adjectives and participles.

The shape for Ammianus must reflect a clear factor of his style, that is, his search for ways to enliven his writing by using words in unusual and emphatic ways.

Of the six shapes for the *Historia Augusta*, the most alike are those for 'Capitolinus', 'Lampridius', 'Pollio' and 'Vopiscus'. 'Spartianus' is generally the same, agreeing with the others particularly for final word-classes. Examination of the scores for the individual lives in Table 4 reveals similar patterns to those found with sentence-lengths, but with some notable differences.

In the 'primary' lives attributed to 'Spartianus', there is a consistently higher percentage of initial nouns than initial verbs, a position reversed in the 'secondary' lives. The same is also true of the 'primary' and 'secondary' lives attributed to 'Capitolinus' and 'Lampridius'. In addition to this, the 'secondary' lives exhibit higher values for initial pronouns and somewhat lower scores for initial adjectives.

The shape for 'Gallicanus' is the most distinctive of the six, but it must be remembered that the *Avidius Cassius*, this writer's only contribution to the collection, is a 'secondary' life. The diagram well reflects the general shape for the other 'secondary' lives: there is the same low proportion of initial nouns to initial verbs, the same high score for initial pronouns and, like the *Aelius*, it has no initial adjectives, and, as in the *Clodius Albinus*, no sentences begin with a participle. Also, in this life, numerals never open a sentence and this is true of the other 'secondary' lives, whereas six out of the eight 'primary' lives have initial numerals of some kind.

For final word-types, the 'primary' and 'secondary' lives are very alike, and no division is readily apparent.

Consideration of the scores for the later lives from the *Macrinus* onwards points to an interesting situation: in these lives, the same tendencies as are evident in the 'primary' lives can be seen, that is, the same higher proportion of initial nouns to verbs (except in four out of sixteen cases), lower scores for initial pronouns and higher values for initial adjectives. Yet, in distributions of sentence-lengths, these later lives were seen to be more akin to the 'secondary' lives, so how does this discovery fit in?

The distributions of sentence-lengths are very much affected by type of composition, and it was asserted above that the lower means for the 'primary' lives are caused by the manner of the major source imposing itself on the composition, while the higher, more widely-spread means for the 'secondary' and later lives are indicative of the personality behind the work coming to the fore. The distributions of initial and final word-types seem to reflect a different process: although form may impose rigid limits on style,

nevertheless, whereas length of sentence may be constrained by these, choice of word-type seems to transcend them. The 'secondary' lives, for which the distributions of sentence-length are the most varied of all the lives, bear clear signs of hasty and clumsy compilation: the high values for initial pronouns indicate something of this process and, on the whole, the usage of word-types is much less varied. These lives are the shortest in the collection and their function seems to be to act as 'fill-ins' between the major lives: it is not *type* but *level* of composition which makes the scores for initial and final word-types in them unlike the rest.

IV

The evidence of the two studies points to one, and only one, conclusion: singular authorship for the *Historia Augusta*. Distributions of sentence-length and of initial and final word-types reveal the homogeneity of the work, while distinguishing it from the control texts studied, which are wholly also distinct from each other. The stylistic variations within the *Historia Augusta* are caused not by there having been more than one author, but by the fact that particular groups of Lives reflect their varied type and level of composition.

Oxford University Computing Service

TABLE 1

TEXT	TOTAL WORDS	TOTAL SENTENCES	MEAN	STANDARD DEVIATION
Spartianus	17,789	1,110	16.03	9.90
Capitolinus	31,593	1,910	16.54	10.60
Gallicanus	2,606	183	14.24	9.64
Lampridius	21,543	1,347	15.99	11.07
Pollio	14,119	850	16.61	11.65
Vopiscus	19,968	1,299	15.37	10.40
H. A. Total	107,618	6,699	16.06	10.67
De Rebus Bellicis	3,347	154	21.73	10.13
Codex	29,699	1,134	26.19	12.83
Ammianus 18	5,043	144	35.02	12.64
Ammianus 19	5,731	160	35.82	13.30
Ammianus 26	6,302	178	35.40	12.83
Ammianus 29	7,793	220	35.42	12.37
Ammianus 30	6,081	165	36.85	10.80
Ammianus 31	9,594	260	36.90	13.24

TABLE 2

	TOTAL WORDS	TOTAL SENTENCES	MEAN	STANDARD DEVIATION
Hadrian	5,106	319	16.01	9.61
Aelius	1,421	71	20.01	13.51
Didius Iulianus	1,585	111	14.28	8.17
Septimius Severus	4,205	280	15.02	8.81
Pescennius Niger	2,264	133	17.02	11.05
Caracalla	1,991	129	15.43	8.99
Geta	1,217	67	18.16	11.12
Antoninus Pius	2,233	150	14.89	9.98
Marcus Aurelius	5,476	339	16.15	9.85
Verus	2,050	114	17.98	11.86
Pertinax	2,576	192	13.42	7.52
Clodius Albinus	2,707	158	17.13	10.75
Opilius Macrinus	2,481	137	18.11	11.08
Maximini Duo	5,405	360	15.01	9.87
Gordiani Tres	5,530	298	18.56	11.57
Maximus & Balbinus	3,135	162	19.35	12.00
(Avidius Cassius	2,606	183	14.24	9.64)
Commodus	3,449	300	11.50	7.98
Diadumenianus	1,660	101	16.44	10.96
Elagabalus	5,779	302	19.14	10.26
Alexander Severus	10,665	644	16.56	11.98
Valeriani Duo	989	68	14.54	11.49
Gallieni	3,619	216	16.75	10.51
Tyranni Triginta	6,558	360	18.22	12.35
Claudius	2,953	206	14.33	11.10
Aurelianus	7,740	482	16.06	10.27
Tacitus	3,072	231	13.30	10.20
Probus	4,110	267	15.39	9.76
Quadrigae	2,312	155	14.92	10.20
Carus	2,734	164	16.67	11.74

