THE NEW STYLOMETRY: A ONE-WORD TEST OF AUTHORSHIP FOR GREEK WRITERS

STYLOMETRY can be defined as the use of numerical methods for the solution of literary problems, most often problems of authorship, integrity, and chronology. As stylometry has been described it seems hardly more than the application of common sense to a literary situation. For example:

> It consists in collecting as many peculiarities of style and grammar as possible from these works [the dialogues of Plato], particularly the *Laws*, which are known, or for good reasons supposed to belong to the author's latest period, and observing the frequency with which these occur in other dialogues. If it is then found, e.g., that one dialogue uses commonly 100 of these, another but 60, it is reasonable to suppose the former to be nearer in time to the *Laws*, i.e. later. Confidence in the soundness of the method is increased by the fact that the dialogues arranged by it show intelligible and self-consistent development of the doctrines expressed.[1]

But the simplicity of this description is deceptive; what features are relevant to a determination of chronology? In what units should they be counted? How large or small a difference between two counts of some feature is to be taken as decisive? If two features seem to offer contradictory evidence, which is to be relied on? No real progress could take place in stylometry until these fundamental questions had been answered. In Greek studies, modern stylometry dates from the work of Dr. W. C. Wake on sentence-length distributions, work which was first published in 1946 and then, more fully, in 1957.[2]

There are three main points of difference between Wake's work and the proposal outlined by Lewis Campbell. Wake concentrates on some event, the construction of a sentence, which happens very frequently. There are two reasons for this concentration on the frequent events. First is the realization that all writers of Greek have much more in common with each other, the language and the conventions of grammar, than they have peculiar to any individual and so what is most easily recognized as characteristic of an individual is not the little which is peculiar to him but the different rate at which he produces features shared with all his colleagues. Second is the knowledge that no safe prediction can be made about the rate of occurrence of a repeated event until five occurrences have been recorded. In other words, to make some judgement about a page of Greek, one must look at features which occur at least five times on an average page. To examine a book of Homer, one must look at events found at least five times in the average book. To examine the smallest possible piece of text, one must look at the most frequent features of Greek texts. In practice any comparison of authors or texts relies on the *difference* between the rates of occurrence of the feature in both texts, and this

[1] H. J. Rose, *A Handbook of Greek Literature* (London, 1964), 262, commenting on the work of Lewis Campbell, Oxford, 1897, and Lutoslawski, London, 1897.

[2] W. C. Wake, 'Sentence Length Distributions of Greek Authors', *J. Roy. Stat. Soc.* Series A, cxx (1957), 331–46.

difference is always less than either rate and so the minimum rate of five occurrences will be much less than is needed to resolve any practical problem.

The second point in which Wake's work differed from the description by Campbell is that he spent much of his time looking at sampling systems and units of measurement. Wake used three systems for taking samples. He numbered all the sentences of some texts and made up random samples by drawing lots using random number tables. While of some interest for mathematical investigations, random samples are of limited application to literary problems. Wake also looked at spread sampling. As the name implies this is used to give a picture of a large text by taking items right through it. If a sample of 200 sentences is needed from a text of 200 pages, then the first sentence on each page might be taken. Not many writers, however, have left enough text to justify spread samples being used. Wake's primary investigation was therefore into block sampling, the method in which a sample starts at any point and carries on until it is large enough to represent the writer. The difficulty confronting this method is that almost all habits of composition are periodic, for example writers tend to group long and short sentences so that in this habit, and for almost all others, there is a minimum sample size which will not prove misleading.

Wake's work has since been extended by the development of a method which enables a whole text to be looked at from the first word to the last and any anomalies within it to be delimited.[1] Measurements have also been made in blocks of twenty successive words marked off right through a text, this to overcome the variable length of the sentence as a unit; and more recently the occurrence of a group of words has been recorded in the spaces between other frequent words, a test completely independent of punctuation.[2]

Modern stylometry deals with the occurrence of features which occur often and records them in whichever unit is best suited to the problem in hand. In this stylometry care is also exercised in the classification of features, for they are defined not only inclusively but exclusively. It is not enough to record the number of nouns found in, say, the first book of Herodotus: the question must be answered, What is the excluded class? Is it all the nouns not found in Book One, or all the words found in Book One which are not nouns? Failure to define classes clearly and unambiguously has been a major restriction on the advance of stylometry.

The third difference between Wake's work and the commonsense proposal lies in his treatment of counts. To Campbell 100 was so much greater than 60 that no question arose from the comparison. But what would he have done if he had three counts, 100, 62, and 60? Would he have taken 60 and 62 as the same? Would he have regarded 100, 68 and 60 differently again? What Wake did was to introduce into classical studies the principles of statistical argument, in particular the dual classification: differences which can be ignored for the good and sufficient reason that they are the product of chance variation, and differences which are too large to be explained by chance and for which some alternative explanation must be offered. In a chronological problem Wake would have looked for features which were present in works written about the

[1] A. Q. Morton, 'The Authorship of Greek Prose', *J. Roy. Stat. Soc.* Series A, cxxviii (1965), 169–224; 'The Integrity of the Pauline Epistles', *Manchester Stat. Soc.*, March 1965.

[2] S. Michaelson and A. Q. Morton, 'Last Words', *New Testament Studies*, xvii (1971).

same time at rates which differed only by chance variation and were so much more frequent, or rarer, in later works that the differences could not be due to chance variation. It was by the introduction of this classification of differences as being due to random variation or being significant and requiring explanation that the transformation of stylometry was effected.

This paper describes a test of authorship using the simplest pattern of choice and so best suited to illustrate the principles used in modern stylometry.

The genesis of this test was the observation, made in a conversation between the authors, that though the occurrence of the pronoun *αὐτός* is irregular, it occurs in bunches in texts and as much as 6,000–10,000 words of text might be needed before stable samples could be obtained, when a computer was printing out a list of occurrences of *αὐτός* with context; no matter how few or how frequent these were in any portion of a text, the proportion of them which were in the genitive case seemed to be constant. The paper describes how this suggestion was made into a test of authorship.

The first point to note is that the classification is precise, two occurrences are recorded: genitive occurrences of *αὐτός* and all cases which are not-genitives.

It is often pointed out that the recording of occurrences in this way ignores the important matter of usage. This is true, but in this context occurrence alone is relevant. Most scholars would want to classify the occurrences of *αὐτός* in terms of usages but this classification is of a different nature. Perhaps the simplest example of the difference between occurrence and usage can be shown by the example with which K. J. Dover begins his book on Greek word order,[1] by quoting a two-word sentence, which Wake would record as such, and then going on to list ten reasons which might explain why the two words are in the order in which they stand. Occurrence is one thing, further classification of occurrence is another.

As has been said, the pattern of occurrence in this example is the simplest possible one, two mutually exclusive events are being counted. This simple pattern means that an equally simple model of how chance would operate in a similar situation can be made; all that is necessary is to visualize a number of cards being used for shuffling and dealing, with the legends 'genitive' and 'not-genitive' written on them in the proportion of one: two.

If we shuffled the cards and drew out sample lots of cards we can illustrate the effects of chance. Drawing one card will tell us very little, for, though the card must be either genitive or not-genitive, no one would suppose on the evidence of one card that all the cards were of one kind. Only when a number of cards have been drawn can we get a picture of the whole group. We would not expect to find in samples of 100 cards drawn from the group 33·3 genitives and 66·7 not-genitives, but we would expect to find numbers close to these estimates fairly often and only occasionally numbers remote from them. We would also notice that the larger the number of cards in the sample, the more closely the sample averages cluster round the group average. The improvement which comes with sample size depends on the square root of the number of items in the sample, so that samples twice as effective as those of 100 occurrences would need to contain 400 occurrences, and another doubling of the efficiency would need samples of 1,600 occurrences.

[1] K. J. Dover, *Greek Word Order* (Cambridge, 1960), 3 f.

The fact that larger samples give better estimates of the rate of occurrence is used to make the first step in the mathematical argument. The argument is more easily understood in an example and Table 1 (A) shows the results of counting the occurrences of *αὐτός* in the first two hundred sentences of each

TABLE 1. *The Occurrence of αὐτός in Thucydides* (O.C.T.)

A. OBSERVED OCCURRENCES

Sample	*Genitive*	*Not-genitive*	*Total*	
Bk. 1	33	58	91	
2	38	60	98	
3	22	65	87	Proportions:
4	24	63	87	Genitives 0·3010
5	24	67	91	Not-Genitives 0·6990
6	38	76	114	
7	22	70	92	
8	35	89	124	
1–8	236	548	784	

B. EXPECTED OCCURRENCES

Sample	*Expected no. of genitives*	*Difference from obs. no.*	*Expected no. of not-genitives*	*Difference from expected no.*
Bk. 1	27·4	5·6	63·6	5·6
2	29·5	8·5	68·5	8·5
3	26·2	4·2	60·8	4·2
4	26·2	2·2	60·8	2·2
5	27·4	3·4	63·6	3·4
6	34·3	3·7	79·7	3·7
7	27·7	5·7	64·3	5·7
8	37·3	2·3	86·7	2·3

C. CHI SQUARED FOR THESE SAMPLES

Bk. 1	1·14	0·49	
2	2·45	1·05	
3	0·67	0·29	Chi squared is
4	0·18	0·08	9·40 for 7 degrees
5	0·42	0·18	of freedom.
6	0·40	0·17	Chi squared is
7	1·17	0·51	9·80 for $p = 0{\cdot}20$
8	0·14	0·06	
Totals	6·57	2·83	

book of the history of Thucydides in the Oxford Classical Text. In the 1,600 sentences there are 784 occurrences of *αὐτός*: 236 of the occurrences, 30·10 per cent, are in the genitive case; 548, 69·90 per cent, are in other cases. We want to know if it is reasonable to assume that Thucydides had a consistent habit in his writing of genitives and the best estimate of his habit is made by pooling all the samples to give an estimate of 30·1 per cent of genitive occurrences. If the hypothesis we have advanced is reasonable, then the differences which are found in each sample between the expected 30·1 per cent genitives and the observed numbers will be explicable by chance alone. If the hypothesis is invalid, then the differences between the expectations and the observations will be larger, or smaller, than the chance variations and the hypothesis must be rejected as untenable.

The next step is to calculate table 1 (B) in which the total number of occurrences of αὐτός in each sample has been divided in the proportion of 30·1 per cent genitives and 69·9 per cent not-genitives. From table 1 (B) it can be seen that for Book 1 the prediction is 27·4 genitives and 33 were counted; in Books 3 and 4, 26·2 genitives were expected, 22 and 24 were counted.

What we want to be able to do is to assess these differences between observations and expectations. The simple figures will not do this, for a difference between observation and expectation of 10 might be quite negligible if the expectation was over a thousand but hardly negligible if the expectation had been six.

One way to assess the differences would be to carry out a card experiment. We could write 'genitive' on 30·1 per cent of the cards and 'not-genitive' on the remainder. If we then shuffled the cards and drew out samples and kept records of the results, we would expect to get, if we drew samples of 91 cards representing Book 1, on the average 27·4 genitive cards or numbers near to this. Rarely would we find numbers far removed from this expectation. By keeping the records of the experiment we would be in a position to say of any size of difference, that it had arisen, by chance alone, once in three drawings, or once in ten drawings, or once in a thousand drawings. In other words we can measure the size of the differences by recording how often each would come about by the operation of chance.

It is not necessary to carry out the card experiment, since statistical mathematics make it possible to calculate, for tables such as table 1 (A) and 1 (B), how often chance would generate the differences between the observations and expectations. One way of doing this is to use the chi squared test. In this test, each pair of figures, an observation with the corresponding expectation, is taken and the difference between them found by subtraction. This difference is then squared, i.e. multiplied by itself, and the result is divided by the expectation. In Table 1 (B) the figures for Book 1 are observed 33, expected 27·4, difference 5·6; squared this is 31·36, divided by 27·4 = 1·14. For the occurrences in other cases in Book 1 the figures are: observed 58, expected 63·6, difference 5·6; squared this is 31·36, divided by 63·6, the result is 0·49 and so chi squared for Book 1 is 1·14 plus 0·49 = 1·63.

To work out how often chance would generate a difference of this size between the observation and expectation we need to know one other thing, the number of degrees of freedom associated with the figure of chi squared. For a table of this kind the number of degrees of freedom is the number of free choices which can be made in dividing up the data which the table comprises. For each sample there is one degree of freedom, for as soon as we decide the number of genitives, or of not-genitives, we fix the other class, which can only be the remainder. Similarly, if we run down the table we can choose different figures in either column for seven of the eight samples, but whatever choices we make in doing this, the eighth sample is fixed: it is the total less the aggregate of the seven previous choices.

In general, for tables of this type, the number of degrees of freedom is found by multiplying one less than the number of columns by one less than the number of rows. For Table 1 the number of degrees of freedom is $(2-1)(8-1) = 7$.

If we then enter tables of chi squared—to be found in most statistical text-books and in all sets of statistical tables—and look at the figures for 7 degrees of freedom, we find the value nearest to the total for Table 1, for which chi squared is 9·40, is chi squared 9·80, $p = 0{\cdot}20$. In these tables p is the probability of the occurrence of the event to which chi squared refers, in this case the occurrence of the set of differences between observations and expectations. p is measured on a scale which runs from $p = 1$ for any event certain to happen at every trial, through $p = 0{\cdot}5$ for an event just as likely to happen as not and so to occur once in two trials, to 0 for an event certain not to happen no matter how many trials are made. In this instance $p = 0{\cdot}20$ tells us that such differences would occur on the average, by the action of chance alone, once in every five trials.

Clearly, if the differences in this habit between the books of Thucydides' *History* are such that chance variation would frequently explain them, then the hypothesis of a consistent habit is well founded. If the differences were to be infrequently produced by chance, some alternative explanation, presumably a change in the habit of writing genitives, would be preferred.

The remaining question is where the line should be drawn between acceptable chance variations and variations not explicable by chance. In general scientific work two *levels of significance*, as the points of decision are called, are in common use. The *5 per cent level of significance*, for which p is 0·05 or less, and the *1 per cent level of significance*, for which p is 0·01 or less. Variations for which p is greater than these figures are called *random sampling differences*, and for these chance is accepted as sufficient explanation. Variations for which p is less than these figures are called *statistically significant differences.*

It may seem to the reader that the lines which have been drawn between the random sampling differences and the statistically significant differences are arbitrary ones. Indeed they are arbitrary, but lines must be drawn somewhere and drawn at a point where the labour involved in making a decision is balanced against the consequences of making a wrong decision. Two other points to bear in mind are that it is not being suggested that a hypothesis should be accepted if $p = 0{\cdot}0501$ and rejected if $p = 0{\cdot}0499$. The figure for p should be stated and placed in one of three ranges. If p is large, one should accept the hypothesis; if p is small, one should reject the hypothesis; if p is intermediate, the indication is that further testing would be desirable.

As soon as further independent tests are employed, the likelihood of making a wrong decision rapidly diminishes. It is true that once in one hundred trials a sample will be rejected on the ground that p is 0·01 or less, although the sample is entirely genuine and the result has been brought about by chance variation, but such a rejection would happen when three independent tests are employed only once in one million trials. Six independent tests would combine to give results significant at the 1 per cent level, by chance only, only once in one million million trials.

A much more practical difficulty is that we cannot ask Thucydides, or any other ancient writer, to write us a few hundred more sentences which might be needed to make a clear-cut decision possible. Even in cases where there is a paucity of evidence, the value of p is a useful indication of the balance of evidence, however narrow it might be.

The pattern of argument has now been sketched out. Samples will be taken from a range of authors and for each of them the hypothesis will be tested that the differences in this habit found within works and between works are only the differences of random sampling for which chance is a sufficient explanation. For comparisons between authors, the differences in this habit will often exceed what random sampling would explain.

A General Survey

At this point it has been shown that the differences in the habit of using genitive occurrences of αὐτός between samples taken from each book of the *History* of Thucydides can be explained by chance variation and that it is therefore reasonable to argue that the samples show the author's habit to be consistent throughout his work.

To transform this suggestive illustration into a test of authorship it is necessary first to show that this hypothesis of a consistent habit is true for writers of Greek in general and then to show that authors differ from each other in this habit enough to enable their works to be distinguished by an examination of the habit. The authors and works to be examined in support of the hypothesis must be chosen with certain factors in mind: that simple writers might differ from complex stylists; that the samples from individual authors must cover long periods of time and wide ranges of subject-matter and include specimens of different genre.

A survey of this kind is not quite as formidable as it sounds, since previous experience has shown that some of these factors have much more effect on habits than others and so attention can be concentrated where significant differences are most likely to arise.

In the course of the review the opportunity will be taken to illustrate two simpler methods of calculating chi squared, one method most useful in the comparison of a pair of samples, the other in the comparison of a number of samples.

Table 2 shows the results of counting the occurrences of αὐτός in a number of authors. The first author in the table is Herodotus. Again the samples are the first two hundred sentences of each book of the history in the Oxford Classical Text. The occurrences are recorded in two columns, genitives and the total of all occurrences. A method of calculating chi squared for such a table is to take the figure of genitive occurrences, 13 for book 1, square this number to get 169, and then divide by the total number of occurrences of αὐτός in Book 1, 39, to give the result = 4·33. The procedure is carried on for all nine samples and the grand total for all nine is 56·92. A similar calculation is made with the totals for all the samples; 152 genitives, square to get 23,104; divide by 415, result 55·67. The difference between the two figures 55·67 and 56·42 is 0·75. To get chi squared from this, one more step is required. The proportion of genitives in all the samples is 152÷415 = 0·366. The proportion of not-genitives is therefore 1·000−0·366 = 0·634. The two proportions multiplied together equals 0·232 and this product is divided into 0·75 to give chi squared as 0·75÷0·232 = 3·23. The number of degrees of freedom is 8 and so p is larger than 0·90. More than nine times in ten trials chance variations between samples with the constant proportion of 36·6 genitive occurrences would be larger than the differences exhibited in the work of Herodotus. The hypothesis of a consistent habit can be reasonably sustained.

The next samples are from Philo Judaeus and are clearly much alike in the occurrence of genitives. For two samples there is a quick way of calculating

[*Turn to p. 98*]

TABLE 2. *The Occurrence of αὐτός in a Selection of Authors*

A

Author and sample HERODOTUS (O.C.T.)	*Number of genitive occurrences* (G)	*Total number of occurrences* (T)	*Calculation of chi squared* (G^2/T)
Bk. 1	13	39	4·33
2	14	46	4·26
3	24	62	9·29
4	18	44	7·36
5	17	49	5·90
6	22	49	9·88
7	12	34	4·24
8	20	55	7·27
9	12	37	3·89
1–9	152	415	56·42
		For 1–9	55·67

Proportions 0·366 0·634
Chi squared = 0·75÷0·232 = 3·28 degrees of freedom 8.
Chi squared is 3·49 for $p = 0{\cdot}90$.

B

Author and sample	*Number of genitive occurrences*	*Number of not-genitive occurrences*	*Total no. of occurrences*
PHILO (Loeb Text)			
On The Creation	30	70	100
The Worse Attacks The Better	33	67	100
Totals	63	137	200

The Calculation of Chi squared.

$G_1 = 30$ $NG_1 = 70$ $T_1 = 100$
$G_2 = 33$ $NG_2 = 67$ $T_2 = 100$
$G_3 = 63$ $NG_3 = 137$ $T_3 = 200$

$$\text{Chi Squared is } \frac{T_3\,(G_1 \times NG_2 - G_2 \times NG_1)^2}{G_3 \times NG_3 \times T_1 \times T_2}$$

$$= \frac{200(2{,}010-2{,}310)^2}{63 \times 137 \times 100 \times 100}$$

$= 0{\cdot}20.$
One degree of freedom.
Therefore $p = 0{\cdot}75$.

C

Author and sample	*Number of genitive occurrences*	*Total number of occurrences*	*Chi squared*
ARISTOTLE (O.C.T.)			
1. *De Caelo*			
A_1–B_1 268ᵃ–284ᵇ	32	100	
2. C_1–D_6 298–313	34	100	5·09; 3 degrees of
3. *Parts of Animals* 639–54	45	100	freedom.
4. *Progression of Animals*—16	32	100	$p = 0{\cdot}15$.
Totals	143	400	

TABLE 2 (C) (*contd.*) *Author and sample*	*Number of genitive occurrences*	*Total number of occurrences*	*Chi squared*
XENOPHON (O.C.T.)			
1. *Hellenica* 6	28	100	
2. *Hellenica* 7	31	100	4·05; 3 degrees of freedom.
3. *Anabasis* 1	19	100	
4. *Anabasis* 3	26	100	$p = 0{\cdot}20$
Totals	104	400	
DIODORUS SICULUS (Loeb Text)			1·13; 1 degree of freedom.
1. Bk. 1 1. 1–38. 9	27	100	
2. Bks. 21, 22, 23	34	100	$p = 0{\cdot}30$
Totals	61	200	
JOSEPHUS (Loeb Text)			
1. *Jewish War*, Bk. 1	24	100	
2. *Jewish War*, Bk. 5	32	100	1·71; 3 degrees of freedom.
3. *Life*, Bk. 1	29	100	
4. *Contra Apion*, Bk. 2	30	100	$p = 0{\cdot}60$
Totals	115	400	
LYSIAS (O.C.T.)			
Oration 1	6	44	
2	22	78	
3	10	49	
4	3	16	
5	1	4	
6	12	60	
7	6	25	
8	4	21	
9	1	5	
10	3	32	
11	4	13	
12	36	107	
13	26	103	
14	24	53	
15	4	15	
16	8	15	Orations 4, 5, 9, 11, 15, 16, 17, 21, 24, 28, 33, and 34 have too few occurrences for the chi squared test. The expected number of genitives would be less than 5.
17	3	5	
18	4	20	
19	11	42	
20	12	33	
21	6	14	
22	12	33	
23	6	24	
24	4	16	
25	8	28	
26	15	49	
27	5	21	Chi squared for the remaining orations is 30·05 for 21 degrees of freedom.
28	4	17	
29	8	27	
30	7	32	
31	17	49	$p = 0{\cdot}10$
32	12	45	
33	3	10	
34	6	13	
35	2	17	
Totals	315	1,135	
Proportion	0·2775		

TABLE 2 (C) (*cont.*) *Author and sample*	*Number of genitive occurrences*	*Total number of occurrences*	*Chi squared*
ISOCRATES (Loeb Text)			
1	6	16	
2	9	34	
3	18	61	
4	41	98	
5	26	88	
6	36	89	
7	19	61	
8	21	76	Chi squared for Works
9	29	79	1–21, 35·74 for 20
10	31	82	degrees of freedom.
11	20	57	$p = 0{\cdot}02$
12	38	110	
13	11	25	For Works 1–15, chi
14	15	50	squared is 21·02 for
15	24	81	14 degrees of freedom.
16	22	56	$p = 0{\cdot}10$
17	13	78	
18	22	73	For Works 16–21, chi
19	20	83	squared is 7·10 for 5
20	4	13	degrees of freedom.
21	4	16	$p = 0{\cdot}25$
Totals	423	1,310	
Proportion	0·3229		
DEMOSTHENES (O.C.T.)			
Oration 1	3	25	Chi squared for Works
2	15	38	1, 2, 3, 4, 8, and 9,
3	6	24	7·89 for 5 degrees of
4	12	35	freedom. $p = 0{\cdot}18$
5	3	12	
6	1	22	If Work 7 is added to
7	6	39	these, chi squared is
8	6	41	13·05 for 6 degrees of freedom.
9	99	41	$p =$ less than 0·05
10	4	46	
Totals	65	323	If Work 10 is added to
Proportion	0·2012		the six, chi squared is 17·24 for 6 degrees of freedom. $p =$ less than 0·01

chi squared, especially if a calculating machine is available. If the occurrences are written in the pattern—

G_1	N_1	T_1	corresponding to	30	70	100
G_2	N_2	T_2	„ „	33	67	100
G_3	N_3	T^{8}	„ „	63	137	200

then chi squared is

$$\frac{T_3(G_1 \,.\, N_2 - G_2 \,.\, N\mathrm{I})^2}{G_3 \,.\, N_3 \,.\, T_1 \,.\, T_2} = \frac{200\ (300)^2}{67 \,.\, 137 \,.\, 100 \,.\, 100} = 0{\cdot}20$$

With one degree of freedom this is equivalent to $p = 0{\cdot}75$. Again there would be no reason to question the hypothesis that the author's habit of using genitive occurrence had not varied from one sample to the other.

There is one point to watch in calculating chi squared for samples such as these. Chi squared is a continuous distribution and can take any value, for example 0·112 or 3·333, but words can occur only in units. With two columns in a table and small numbers of occurrences it can be important to allow for the difference produced by the steps of one unit in which words occur and the continuously variable chi squared. This is done by reducing the difference between observations and expectations by one half, 0·5, before squaring the differences in the calculation of chi squared. This correction is known by the name of its originator, Yates. For the rapid method just illustrated the correction is made by subtracting half of T_3 from the value of $(G_1 . N_2 - G_2 . N_1)$ before squaring. In that case the result is to make the difference 250 and not 300 and so chi squared is 0·14 rather than 0·20.

The next four sets of samples are taken from the works of Aristotle, Xenophon, Diodorus Siculus, and Josephus.

In the Aristotelian samples the only comment to be made is that since the pronoun occurs rarely, the samples are about 700 sentences for each hundred occurrences of *αὐτός* and the test must be correspondingly insensitive. For the four samples p is 0·15 and the differences between the samples are not statistically significant.

The two sets of samples from Xenophon and Diodorus Siculus show no statistically significant differences in the occurrence of genitives. The Josephus samples include works written by him both before and after his decision to use a secretary to improve his Greek. Whatever changes the secretary made in Josephus' *style*, he did not alter his rate of using genitive occurrences of *αὐτός* sufficiently to create statistically significant differences in this habit.

The next sets of samples are the most important in the general survey because in all previous examinations the orators have shown most variation in stylistic habits, and so the samples are from the works of Lysias, Demosthenes, and Isocrates.

The first of these sets is all 35 works of the *Corpus Lysiacum*. No less than 13 of these orations have so few occurrences of *αὐτός* in them that the expectation of having five occurrences of genitives is not reached. For the 22 orations which can be safely tested, chi squared is 30·05 for 21 degrees of freedom, p is 0·10; no statistically significant differences are found between the orations and there is no obstacle to the hypothesis that all have been written with a constant habit of using genitives.

Looking at the individual orations, it is to be expected that one in a set of 22 would show differences significant at the 5 per cent level. One oration does show such differences, oration 14, and it is of this oration that Dover has written that it contains 'an exceptionally high proportion of postponed *αὐτός*'.[1]

The next set of samples is the first ten orations of Demosthenes. Of the ten, two are acknowledged to be spurious, and orations 5 and 6 are too short to provide the minimum number of genitives for safe predictions. For the other six orations, chi squared is 7·79 for 5 degrees of freedom and p is about 0·18. If oration 7 is added to the six others, then chi squared rises to 13·05 for 6 degrees

[1] K. J. Dover, *Lysias and the Corpus Lysiacum* (University of California, 1968), 137 f.

of freedom and the differences become statistically significant at the 5 per cent level. If oration 10 is added to the set of six, chi squared is then 17·24 for 6 degrees of freedom and p is less than 0·01. The two spurious orations are shown to be such.[1]

The final set of samples is the 21 works of Isocrates. The results of applying this test to the corpus are in agreement with other tests. If all 21 works are included the differences are statistically significant, but if either the forensic orations or the epideictic orations are excluded the statistically significant differences disappear. As the forensic orations are his early works and the epideictic orations are, on the whole, his later works, it is not possible to say whether the differences are due to the contrast in genre, the long interval of time, over 40 years, or the combination of both.

The general review has shown that for a number of authors the habit of writing a constant proportion of genitives among the occurrences of αὐτός is supported by almost all the examples. Only the combination of the long periods of time and the range of genre in Isocrates produces statistically significant differences in the habit. The habit varies from author to author but not greatly; it has the same mathematical form as the habit of using the particle δέ as the second or third word of sentences, a habit in which some writers have a proportion as low as 3 per cent and as high as 67 per cent. The habit of writing genitive occurrences of αὐτός varies much less from one author to another. It is not to be neglected for that reason alone and should be included in any routine of investigation.

Two Cases of Disputed Authorship

1. *Plato's Seventh Epistle*

In Table 3(A) are shown the data for three Platonic samples. The first is the *Apology*, an early work; the second is a sample from Book 1 of the *Laws*, a late work, and the sample was continued until it had the same number of occurrences of αὐτός in it as the disputed *Seventh Epistle*. The proportions of genitives in the *Apology* and the *Laws* are almost identical and a comparison of the *Laws* sample with *Epistle* 7 gives chi squared = 1·45 for 1 degree of freedom and so $p = 0{\cdot}25$.

On the face of it, all three samples are homogeneous. But it is the *integrity* of *Epistle* 7 that is questioned,[2] and inspection of the sequence of occurrences shows that between 314 b and 344 b of the text, there are only 2 genitive occurrences in 31. This portion of the text therefore differs from the rest of the epistle by an amount which is statistically significant at the 5 per cent level, and the effect is unlikely to be due to any periodic variations since sequences of 31 observations from the other samples show no such significant differences.

2. *The Pauline Epistles*

Table 3(B) shows the results of counting the occurrences of αὐτός in all 14 epistles of the Pauline corpus in the United Bible Societies' Greek New Testament, edited by Aland, Black, Metzger, and Wikgren, London and New York, 1967. Previous examination of these epistles[3] had shown that if Paul is defined

[1] F. Blass, *Die Attische Beredsamkeit*, iii (1). 55.

[2] M. Levison, A. Q. Morton, and A. D. Winspear, 'The Seventh Letter of Plato', *Mind* xxvii (1968), 309–25.

[3] Michaelson and Morton, loc. cit.

TABLE 3

Author and sample	*Number of genitive occurrences of αὐτός*	*Total number of occurrences of αὐτός*	*Chi squared*
A			
PLATO (O.C.T.)			
Apology	30	98	
Laws, Book 1	43	141	
Epistle 7	34	141	
B			
THE PAULINE CORPUS			
1st Corinthians	37	84	
Galatians	8	26	1·45
Romans	79	156	
2nd Corinthians	28	61	
Ephesians	39	66	5·47
Philippians	8	30	2·03
Colossians	16	45	0·38
1st Thessalonians	7	24	
2nd Thessalonians	5	17	
1st Timothy	1	6	
2nd Timothy	11	16	
Titus	5	8	
Philemon	—	3	
Hebrews	70	140	1·77

The figures for chi squared are for a comparison with the sum of *1st Corinthians* and *Galatians*, except for the first comparison of *1st Corinthians* with *Galatians*

to be the author of *Galatians*, then only *1st Corinthians* and *Galatians* are purely Pauline. The anomalies in *Romans* and *2nd Corinthians* are small and are unlikely to show up in this test.

A comparison of *1st Corinthians* and *Galatians* gives chi squared 1·45 for 1 degree of freedom, $p = 0{\cdot}25$. When the two samples are added to give an estimate of the author's habit and compared with *Romans* and *2nd Corinthians*, chi squared is 0·71 and 0·40 for 1 degree of freedom; for *Romans* $p = 0{\cdot}40$, for *2nd Corinthians* $p = 0{\cdot}60$.

Before the shorter epistles can be examined, the minimum sample size must be established. If *1st Corinthians* is divided into halves and these are compared, chi squared is 2·37 for 1 degree of freedom, $p = 0{\cdot}15$. If the epistle is divided into three and the parts compared, chi squared is 2·68 for two degrees of freedom and $p = 0{\cdot}25$. If the epistle is divided into quarters and these are compared, chi squared is 22·8 for 3 degrees of freedom and p is less than 0·001. If *Galatians* is added to this comparison of thirds, then chi squared is 4·21 for 3 degrees of freedom, $p = 0{\cdot}25$. The conclusion is that samples of 26 to 28 observations are adequate to represent this author but anything less is liable to be misleading because of the grouping of the two classes of occurrences.

When *Ephesians*, *Philippians*, and *Colossians* are compared, chi squared is 11·00 for 2 degrees of freedom, p is less than 0·01 and the differences are significant. When compared with the *1st Corinthians–Galatians* expectation, only *Ephesians* gives a statistically significant result.

The only other epistle long enough to make a safe comparison is *Hebrews* and for this epistle chi squared is 1·77 for 1 degree of freedom and so $p = 0{\cdot}15$. The only comment to be offered is that, as in almost all other tests, *Hebrews*, widely rejected as not Pauline, much more closely resembles the genuine epistles than does *Ephesians*, still generally accepted.

Conclusion

A simple binomial classification of this kind cannot be the basis of a sensitive test of authorship, especially when the differences between authors in the habit are not great. But the test is a useful addition to those already established and it has the advantage that it can be readily incorporated in a computer programme.

University of Edinburgh S. Michaelson and A. Q. Morton