A Multivariate Analysis of Latin Elegiac Verse

Robert E. Blackith

Language and Speech 1963 6: 196
DOI: 10.1177/002383096300600402

A MULTIVARIATE ANALYSIS OF LATIN ELEGIAC VERSE

ROBERT E. BLACKITH
Imperial College, London

Multivariate analyses offer the possibility of distinguishing the various independent components of style which contrast one sample of text with another. The analysis will show which of the stylistic components is of the greatest importance in any particular contrast. The application of these techniques to the greater part of the extant corpus of Latin elegiac verse shows that the concept of style in this setting can be dissected into at least three independent components. Two of these components, associated previously with the entropy and elision frequency of the text, are under more or less conscious control by the author, but the third, associated previously with the mean syllable number per word, reflects the date at which the verse was written and is apparently refractive to counterfeiting. Severe mental stress in the author is reflected in this component.

INTRODUCTION

There are obvious possibilities of making literary analyses more objective by the use of statistical techniques and several attributes of style have been used for this purpose. These attributes include the mean number of letters per word (Meldenhall, quoted by Williams, 1956), sentence length (Yule, 1939 : Williams, 1956), syllable frequency distributions (Fucks, 1952) and various aspects of the entropy of the text, or its degree of looseness of construction with respect to several possible attributes, for example the syllable number distribution (Herdan, 1956 : Fucks, 1952 ; Whatmough, 1956). Cox and Brandwood (1959) were the first to use formally multivariate analysis in order to discriminate between the frequencies of clausulae in Plato's works so as to be able to rank these works in the order in which they were written. However, although the structure of their analysis is multivariate, the discrimination is carried out along a single vector, and the analysis is effectively multivariable rather than multivariate. Fucks, on the other hand, was the first author to reveal the composite nature of literary style, uncovering a vector which distinguished languages and one which distinguished authors within any one language. Thus in his analysis, which was formally univariate although multivariate in conception, the distinction between the styles of Aldous Huxley and Galsworthy in English was found to be parallel to that between Thomas Mann and Rilke in German.

There is, therefore, an opportunity afforded by multivariate analysis, of bringing to light the different kinds of distinction which underlie the comparisons of style. Such a situation makes use of the greater generality and power that comes from the transition to vector techniques as compared with the univariate, scalar methods. This transition to vector and tensor techniques of analysis has proved fruitful in several branches of biology and psychology (Blackith, 1961) and may be expected to serve the same ends in linguistic analyses. Essentially, this benefit stems from the possibility of making qualitative differences into quantitative ones, by using the

orientation of the vectors (linear compounds of the attributes used) to represent the nature, as the associated generalised distance represents the amount, of the distinction between any two samples. Since the angle between the directions of the several vectors can readily be computed, a direct measure of similarity of stylistic contrasts is available.

These applications of multivariate analysis offer the advantage that no *a priori* hypotheses about the nature of literary style are required : the number of independent patterns of thought exhibited in the works of any one author may be ascertained objectively by factorising the matrix of correlations (or covariances) of the attributes which have been measured. In most cases the nature, as well as the number, of these patterns will become evident from the way in which highly characteristic and readily identifiable passages of prose or verse orient themselves along the various axes.

In this pilot study a limited range of material, all Latin elegiac verse, and an even more limited number of characters, is analysed as much to suggest further possibilities in larger investigations as to solve problems particular to this verse.

Literary material and methods of analysis

The elegiac verse (distychs composed of a hexameter followed by a pentameter) studied in this work comprises the substantially greater part of the extant poems of Ovid, Catullus, Tibullus, Propertius, Martial, Avianus, Rutilius, Namatianus, Pentadius and Lactantius in this metre. For comparison, some poems of the 20th century Latin elegiast Henrici Paoli have also been analysed. Loeb or Oxford editions of the classical authors were used, the poems of Paoli being taken from an edition of the 'Isargi Carmina' published in 'Vita Latina' for January, 1958, by Editions Aubanel, Avignon, France.

The attributes of the verse chosen for study were the elision frequency (number of elisions per 1,000 words), the mean syllable number per word, the variance of the distribution of the syllable number, and the entropy of mixing of the words of different syllable numbers. Elision frequency, often used as a criterion of style, is usually recorded as the number of elisions per 100 lines (Platnauer, 1951). The opportunities for elision are, however, at word endings rather than a function of the number of lines, and this fact is reflected in the significantly greater number of words in lines *with* elision(s) than in lines *without* elision. In the body of elegiac verse studied here there are 28 lines containing 10 or more words *with* elisions but only 15 such lines *without* elisions. The trace of the matrix of distances between words of the same syllable number was used to good effect by Fucks, but in the present context was found to be very unstable, and weighted heavily by the occurrence of pairs of rare, highly polysyllablic words. This attribute was therefore abandoned. The elision frequency has been known to follow the dictates of stylistic fashion since antiquity, and individual authors have managed to hold their own characteristic elision frequency remarkably constant, probably quite consciously so. Two samples (see Table 1) from Tibullus have elision frequencies of 19.1 and 21.4, two from Rutilius Namatianus

TABLE 1

AUTHOR	WORKS		NO. OF WORDS IN SAMPLE	ELISION FREQUENCY PER 1,000 WORDS (E)	MEAN SYLLABLE NUMBER (i)	ENTROPY (S)	VARIANCE (V)
OVID	Heriodes Bks.	I & II	1658	19.90	2.259	0.564	0.851
	,,	III & IV	2106	20.41	2.219	0.550	0.792
	,,	V & VI	1962	14.78	2.265	0.558	0.818
	,,	VII, VIII & IX	3044	24.31	2.249	0.558	0.827
	,,	XV	1405	24.19	2.229	0.565	0.807
	,,	XVII & XVIII	2818	15.96	2.147	0.539	0.754
	,,	XX	1603	12.47	2.126	0.549	0.801
	,,	XXI	1629	18.30	2.133	0.539	0.761
	Amores	I ; i-vii	2242	15.16	2.225	0.552	0.793
	,,	I ; viii-xiii	2046	21.50	2.234	0.545	0.763
	,,	II ; i-ix	2578	27.15	2.203	0.548	0.778
	,,	II ; x-xix	2582	20.10	2.211	0.549	0.795
	,,	III ; i-vii	2997	19.68	2.233	0.557	0.831
	,,	III ; viii-xv	2420	16.94	2.264	0.556	0.814
	Ars amatoria	I ; 1-398	2515	17.49	2.231	0.553	0.800
	,,	I ; 399-772	2356	24.61	2.249	0.557	0.811
	,,	II ; 1-372	2375	17.26	2.234	0.554	0.814
	,,	II ; 372-746	2402	20.39	2.201	0.550	0.753
	,,	III ; 1-432	2730	18.31	2.235	0.556	0.819
	,,	III ; 433-812	2427	18.12	2.196	0.540	0.749
	Remedia amoris	I ; 1-396	2562	22.24	2.190	0.545	0.768
	,,	I ; 397-814	2679	19.78	2.207	0.555	0.816
	Fastii	III ; 1-403	2481	9.67	2.284	0.549	0.778
	,,	III ; 404-884	3042	23.33	2.266	0.556	0.806
	Tristia	I ; i-iv	2352	17.43	2.207	0.558	0.827
	,,	I ; v-xi	2249	23.12	2.237	0.576	0.915
	,,	II ;	3633	26.14	2.248	0.568	0.875
	,,	III ; i-vi	2409	28.22	2.183	0.557	0.835
	,,	III ; vii-xiv	2677	18.30	2.200	0.552	0.809
	,,	IV ; i-v	2267	20.73	2.186	0.554	0.826
	,,	IV ; v-x	2092	18.64	2.204	0.566	0.884
	,,	V ; i-vi	2438	22.96	2.183	0.592	0.882
	,,	V ; vii-xiv	2458	19.52	2.143	0.548	0.798

	ex. Ponto	I ;	i-iv	2439	19.68	2.203	0.561	0.844
	„	I ;	v-x	2527	16.22	2.066	0.547	0.796
	„	II ;	i-v	2631	21.28	2.204	0.562	0.844
	„	II ;	vi-xi	2344	16.63	2.166	0.543	0.771
	„	III ;	i-iii	2502	22.38	2.163	0.554	0.788
	„	III ;	iv-ix	2283	23.21	2.177	0.554	0.819
	„	IV ;	i-viii	2825	15.92	2.176	0.546	0.782
	„	IV ;	ix-xvi	3290	20.36	2.179	0.558	0.838
	De Medic. Fac. Lib.		612	612	16.33	2.312	0.563	0.840
	Nux			1183	30.43	2.175	0.552	0.820
	Ibis		1-316	1998	21.52	2.239	0.562	0.866
	„		317-642	2009	11.94	2.302	0.588	0.995
ANONYMOUS	Cons. ad Liviam		1-270	1737	48.93	2.211	0.556	0.818
	„		271-474	1279	35.18	2.240	0.568	0.868
	Eleg. ad Maecenatem			1104	9.05	2.280	0.558	0.820
PROPERTIUS	Elegiae	II ;	i-x	2238	45.57	2.184	0.557	0.825
	„	II ;	x-xx	2197	44.60	2.172	0.546	0.758
	„	II ;	xxi-xxviiiB	2343	44.81	2.112	0.539	0.757
	„	II ;	xxix-xxxivB	1965	40.71	2.159	0.546	0.781
	„	III ;	i-xii	3441	41.55	2.212	0.552	0.793
	„	III ;	xiii-xxv	2833	47.29	2.214	0.554	0.806
TIBULLUS	Delia	I		5225	19.10	2.183	0.543	0.768
	Nemesis	II		2807	21.37	2.216	0.542	0.763
TIBULLUS/LYGDAMUS		III		1778	24.18	2.243	0.544	0.756
CATULLUS			lxv-lxx	2061	61.64	2.235	0.573	0.884
			lxxi-cxvi	1971	107.61	2.128	0.566	0.890
RUTILIUS NAMATIANUS		I ;	1-370	2074	17.87	2.470	0.578	0.879
		I & II ;	317-644	1810	17.12	2.498	0.584	0.969
LACTANTIUS	Phoenix			1064	15.03	2.197	0.552	0.800
MARTIAL	Epigrammata up to xl of I			2536	13.80	2.228	0.570	0.883
	Epigrammata up to xcviii of I			1961	16.82	2.170	0.558	0.859
AVIANUS	Fabulae		i-xxi	1915	19.32	2.339	0.565	0.851
	„		xxii-xlii	1945	15.93	2.344	0.568	0.828
HENRICI PAOLI	Carmina Isargi			1876	30.91	2.298	0.537	0.719
PENTADIUS				499	46.09	2.214	0.521	0.686

have 16.8 and 17.1, four from Propertius 45.6, 44.6, 44.8 and 40.7. Each of the samples taken amounts to some 2,000 words, and in most instances it proved expedient to divide each poem into two parts, since poems often run to some 5,000 words.

Where elision occurred in the scansion of a line, the effective, rather than the apparent syllable number for the affected word was recorded. The entropy function was computed from the expression

$$S = -k \sum p_i . \log p_i \quad . \quad . \quad . \quad . \quad . \quad (9)$$

of Fucks (1952) and may be regarded as a measure of the degree of organisation of the text with respect to words of different syllable number.

Statistical analysis

The original intention was to perform an analysis along canonical axes, in which case the variation within poems would have been assessed by contrasting the two halves into which most poems were divided, and the variation between poems would have been computed directly from the total values for each attribute in any one poem. However, substantial cleavages were found in the course of a preliminary examination of the split poems during which the mean syllable number and its variance were assessed against their own internal measures of error. Although no systematic trends could be seen, the variation from one half of a poem to the other was often statistically significant and seemed to be at least as great as that from one poem to another by the same author. For this reason the canonical analysis was abandoned and a principal component analysis carried out on the whole body of evidence, treating the half-poems as individual entities (for details of the statistical procedures, see Rao, 1952).

As a first step, the correlation matrix showing the inter-relationships between the four attributes over the whole field of Latin elegiacs was prepared. As Table 2 shows, the entropy and variance are closely correlated, and probably represent slightly different ways of measuring essentially the same feature of the text, its degree of organisation with respect to words of different syllable number. A surprising feature is the low correlation between the elision frequency and the remaining attributes. The negative values for the correlation coefficients probably do not bear the implication which one might attribute to them at first sight, e.g., that the number of elisions decreases as the syllable number increases. Elision frequency increased and syllable number decreased during the decline of Latin verse from the first century to the twentieth, and both changes are probably associated with general trends in time which render any correlation between them essentially spurious.

The effective independence of elision frequency as a stylistic attribute, and the close association between entropy and variance, suggest that there is an underlying structure of these elegiac poems which multivariate analysis ought to be capable of laying bare. The latent vectors of the characteristic equation of the correlation matrix may be used as sign-posts to indicate the general outlines of this latent structure. The characteristic equation is

$$| A - \lambda I | = 0$$

TABLE 2

THE CORRELATION MATRIX

E	1.00	−0.24	−0.03	0.01	(Elision frequency)
i		1.00	0.49	0.42	(Mean syllable number)
S			1.00	0.93	(Entropy of organization)
V				1.00	(Variance of syllable number distribution)

TABLE 3

LATENT ROOTS AND VECTORS OF THE MATRIX

Root	Vector			
2.272	0.100 E	−0.464 i	−0.631 S	−0.613 V
1.091	−0.889 E	+0.352 i	−0.176 S	−0.230 V
0.571	−0.445 E	−0.810 i	−0.222 S	+0.311 V

where A is the correlation matrix, I is the unit matrix of the same rank, and the λ's represent the latent roots of the determinantal equation. Whilst there are necessarily as many roots as there are measured attributes, not all of these roots need generate latent vectors accounting for statistically significant components of variation.

The four latent roots and their associated vectors (or principal components) were evaluated on a Ferranti 'Pegasus' electronic computer, and are shown in Table 3. It is evident that the last latent vector takes up a negligible amount (about 2%) of the total variation.

RESULTS AND DISCUSSION

The first latent vector

The greater part (57%) of the variation of style between the samples of words investigated lies in their fluctuating degree of organisation. This quantity is measured by the first of the latent vectors, heavily weighted on the entropy and variance, although the mean syllable number per word also plays its part in determining the orientation of the vector. The numerical scores for each sample of words correspond to their position along each of the three meaningful axes of variation, which are necessarily orthogonal. Thus large negative values along the first vector represent relatively disorganised poetry, large positive ones representing more closely knit text. It is noteworthy that the Golden Age poets (especially Propertius and Tibullus) wrote

closely organised texts, later authors, notably Martial, Avianus and Rutilius Namatianus relaxing their organisation in this respect. Catullus, writing before the Golden Age of Latin verse, also scores low in terms of the criteriâ used to judge textual organisation, so that the apogee of Roman culture can in this measure be traced numerically. With care and a flair for the Golden Age style, however, some late authors have refined their verse to conform in this respect to the canons of Augustan verse, as witness the high scores of Pentadius' fragment and the elegiacs of Paoli.

It is evident that the degree of organisation of the text is to some extent conditioned by the stresses of the author's environment. Within the substantial *corpus Ovidianum,* systematic trends are to be found. Apart from the very first poem analysed (the earliest Heroides) Ovid's writings from his happier years at Rome score uniformly well by this criterion, the 23 poems or half-poems ranging in score from −0.49 to 2.07. However, the early poems from exile at Tomis on the Black Sea coast (now Constanza) score altogether lower, the two halves of the diatribe 'Ibis' (which was written against some unspecified person whom Ovid deemed to have tried to defraud him of his Roman possessions in the turmoil of banishment) scores −1.08 and −4.45. The first half of the first of the Tristia, known to have been written during the voyage from Rome to Tomis at a time when Ovid found the emotional adjustment to banishment almost unbearable, scores −0.21, but the second half, written when the full horror of exile could no longer be evaded, scores −2.36. The next poem, the first half of the second Tristia, scores almost as badly at −1.52, but thereafter there is an erratic but general improvement corresponding to the poet's adaptation to his surroundings. It is evident from the contents of the 'Tristia' and of the 'Epistulae ex Ponto' that Ovid became at least inured to life at Tomis, and that he even derived a measure of contentment from the regard in which he was held by the inhabitants. The last of the Ovidean poems have a texture, as measured by the first of the principal components in this analysis, which is not appreciably inferior to that of much of the pre-banishment epoch. Fraenkel (1945) has noted how Ovid's 'resilient art was again on the upsurge when death overtook him'.

The second latent vector

This vector takes up some 28% of the total variation between the samples. It is essentially a reflection of the differential control of elision frequency by the various authors, to which allusion has already been made. Being, so far as one can tell, under conscious control at least in so far as elision frequency is an important element of the literary style peculiar to each author, the evidence afforded by the elision frequency is less reliable than that of attributes not so much consciously influenced by the vagaries of taste, or the desire to write pastiche or even forgeries.

Nevertheless, where the position of a poem along this second vector is incompatible with its putative authorship the evidence points to the fact that the supposed author did not in fact write the poem. Such is the case with the 'Consolatio ad Liviam', variously attributed to Ovid and to his friend Albinovanus Pedo, together with the 'Elegiae in Maecenatem', composite poems in the same consolatory vein. The compounding of the elision frequency into a discriminatory vector sharpens the capacity of

this attribute to detect the discrepancies inherent in these poems beyond that of the elision frequency alone, whose disparity in the two poems has been noted by Duff (1935). It is in fact virtually impossible for the 'Consolatio ad Liviam' to have been written by Ovid, since the positions along the second latent vector for the two halves of the poem are -1.55 and -0.95, each of which falls short of the mean score for genuine Ovidian poems (41 samples) by approximately 3 and 2 standard errors respectively, if one may be permitted to make use of an oriented multivariate significance test in this way. There is evidently a need for significance tests directed along a particular vector rather than the usual empirical orientation of such tests as Wilks' criterion (Rao, 1952) which are irrelevant to the present purpose. The 'ad Maecenam' elegy, on the contrary, scores highly on this vector, exceeding all but one of the scores of genuine Ovidian samples. This circumstance makes the common authorship of the two poems intrinsically improbable, notwithstanding the fact that there is in the elegy a claim to have already written the consolation. If the same hand wrote both poems, it is unstable to a degree unthinkable in a Golden Age poet. Only Catullus, writing at a time when elegiac verse was still experimental, has as wide a range of a scores along this axis of variation. Wilkinson (1955) described the 'Consolatio' as a certain forgery, and it is thus probable that the Elegy 'ad Maecenam' is a forgery also, rather than a poem whose authorship has been forgotten.

The poem to a nut tree 'Nux' is, however, consistent with Ovid's delicate workmanship in all respects that can be settled by this statistical investigation. In view of the extreme difficulty of forging so delicate a web of stylistic features as Ovid could call upon and yet meeting the statistical criteria, this finding greatly strengthens Ganzenmüller's opinion (Ganzenmüller, 1910) that the poem is indeed Ovidian.

The third latent vector

This vector accounts for just over 14% of the total variability. The relatively modest contribution in this respect does not do justice to its value as a discriminant of Latin elegiac verse, since it is perhaps the most rewarding of the three vectors to explore. In this discriminant the main attributes are the mean syllable number and the elision frequency, which on this occasion vary together rather than, as in the first vector, in opposite directions. Of the two, the syllable number is substantially the more important attribute.

The outstanding merit of this linear compound of the four attributes is its ability to predict the date of the poem in a way which is apparently refractive to counterfeiting. Table 4 shows the close association between the probable dates of the poets whose work has been analysed here and the scores of their poems along this third axis of variation, independent of the two axes already discussed. The score rises to a maximum in Augustan times then drops steadily to the present day, although the fall is naturally faster while the stylistic features of Latin elegiacs were evolving during the first centuries of this era than during the long, static, period from the decline of Rome to the present day.

Within the *corpus Ovidianum,* however, there is a marked trend. In particular, there is a significant change when the pre-banishment poems are compared with those from

TABLE 4

Variation of the score on the 3rd latent vector with time

AUTHOR	DATE (FLOREAT)	MEAN SCORE ON 3RD VARIATE
Catullus	60 B.C.	−0.70
Propertius	30 B.C.	−0.32
Tibullus	25 B.C.	−0.16
Ovid (pre-exile)	5 A.D.	−0.03
Martial	90 A.D.	1.02
Lactantius	300 A.D.	0.41
Pentadius	300 A.D. (?)	−1.99
Avianus	400 A.D. (?)	−0.88
Rutilius Namatianus	420 A.D.	−1.79
Henrici Paoli	1930 A.D.	−2.05

Tomis. Prior to his banishment Ovid wrote poems with a mean score of −0.0261 on this third latent vector, whereas afterwards his mean score was 0.7034. Again using a primitive form of oriented multivariate significance test, the difference between the two scores exceeds the standard error of either by a factor of 2 or more. It is tempting to speculate on the causes of these changes. Ovid himself complained during his banishment that no one at Tomis spoke Latin and few had any Greek, so that he was forced to learn the two local languages. Such total isolation, over some eight or nine years, can hardly have failed to leave its traces on the style of even so practised and brilliant an elegiast as Ovid. The change to a higher score follows, however, immediately on the banishment (i.e. with the first of the Tristia) and seems to be no greater at the end of Ovid's life than in his first days of exile. It seems reasonable, therefore, to attribute to mental stress, rather than to isolation as such, the change in style : in this connection we have to remember that Ovid's life at Tomis was not merely sad in the way that any involuntary exile's life is sad, but also a constant round of physical hardship and danger in a primitive and inclement region.

As a postscript to this investigation, the limited capacity of the four attributes chosen for this work to illustrate stylistic changes must be emphasized. Even this superficial investigation, however, necessitated the scansion of some 100,000 words. There is evidently a mine of information about the components of 'style' to be obtained from a more detailed and comprehensive study, even within the field of classical studies. As yet the influence of metre on a poet's style is almost wholly unknown as are the constraints of language in the case of bilingual poets, to name only two obvious subjects for further research.

References

BLACKITH, R. E. (1961). Multivariate statistical methods in human biology. *Medizinische Dokumentation*, 5, 26.

COX, D. R. and BRANDWOOD, L. (1959). On a discriminatory problem connected with the works of Plato. *J. R. Stat. Soc., B*, 21, 195.

DUFF, J. W. and DUFF, A. M. (1935). Minor Latin Poets (London).

FRÄNKEL, H. (1945). Ovid ; A Poet Between Two Worlds (Berkeley, Calif.).

FUCKS, W. (1952). On mathematical analysis of style. *Biometrika*, 39, 122.

GANZENMÜLLER, C. (1910). Die Elegie Nux und ihr Verfasser (Leipzig).

HERDAN, G. (1956). Language as Choice and Chance (Groningen).

PLATNAUER, M. (1951). Latin Elegiac Verse (Cambridge).

RAO, C. R. (1952). Advanced Statistical Methods in Biometric Research (New York).

WHATMOUGH, J. (1956). Poetic, Scientific and Other Forms of Discourse (Berkeley, Calif.).

WILKINSON, L. P. (1955). Ovid Recalled (Cambridge).

WILLIAMS, C. B. (1956). Studies in the history of probability and statistics, IV: A note on an early statistical study of literary style. *Biometrika*, 43, 248.

YULE, G. U. (1939). On sentence length as a statistical characteristic of style in prose. *Biometrika*, 30, 363.