'An Ocean Where Each Kind . . .': Statistical Analysis and some Major Determinants of Literary Style

J. F. Burrows[1]

Department of English, The University of Newcastle, New South Wales, 2308, Australia

Abstract: The statistical analysis of literary texts has yielded valuable results, not least when it has treated of the frequency patterns of very common words. But, whereas particular frequency patterns have usually been examined as discrete phenomena, it is possible to correlate the frequency profiles of all the very common words, to subject the resulting correlation matrix to eigen analysis, and to present the results in graphic form. The specimens offered here deal, first, with differences among Jane Austen's characters and, secondly, with differences between authors. The most striking general differences among the authors studied relate to historical eras and authorial gender.

Key Words: literary criticism, stylistic analysis, literary statistics.

The evidence offered in this paper both extends and generalizes the argument of my book, *Computation into Criticism: A Study of Jane Austen's Novels and an Experiment in Method* (Oxford, Clarendon Press, 1987). The argument is extended by taking specimen-pieces from many more novels than were treated there. It is generalized by gathering Jane Austen's characters into appropriate groups and allowing the pieces by the other novelists to group themselves as they may. By these means, I hope to contribute to the resolution of a celebrated paradox: that many recent studies in "stylistics" and "stylometrics" arrive at intuitively satisfying results while relying on methods that have been condemned as circular, selective, and statistically impoverished.[2] As I shall show, even the most common word-types of our language make up a closely patterned web of meaning: since that is so, the "stylistician" who examines any of its brighter filaments is likely to arrive at more valuable findings than his method might seem to promise. The "redundancy" defined by communication-theorists is his ultimate support: if English bears its meanings with a high enough degree of "redundancy", even a crude instrument will receive its more important "messages" and enable a skilled interpreter to draw valuable conclusions. The method of analysis employed in this paper complements those of stylistics: it may not lead to conclusions as subtle as those can be: but it is free of circularity; it is selective only in a limited sense; and it is statistically robust.

Each of my first four graphs is based on an analysis of some 40% of all the word-tokens (comprising the thirty most common word-types) that make up Jane Austen's dialogue. Each graph thus incorporates the evidence of more than

---

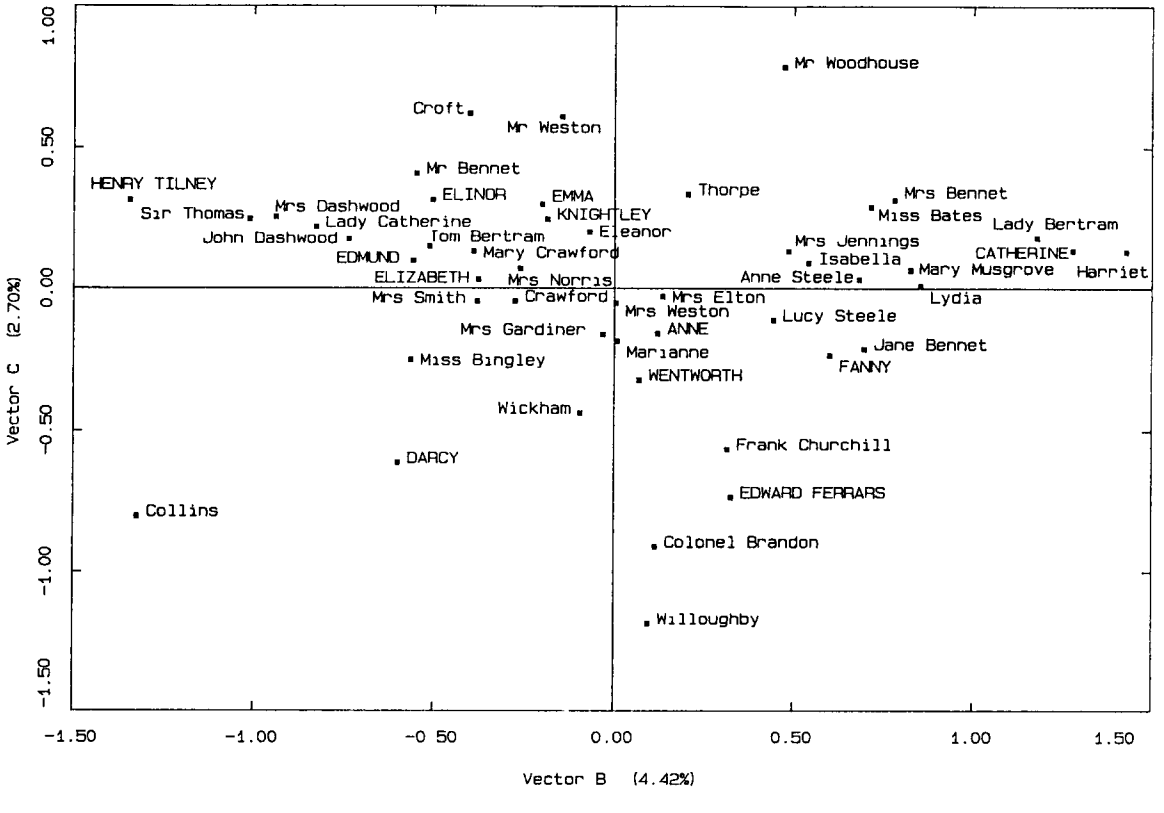
John Burrows has been Professor of English at the University of Newcastle, N. S. W. since 1976. He was Commonwealth Fellow of St. John's College, Cambridge in 1979—80 and a Visiting Research Fellow at the Institute for Advanced Studies in the Humanities at the University of Edinburgh in 1988. His earlier publications are mostly in the field of Australian literature. His chief research interest since 1979 has been in the computer-assisted analysis of literary texts. Publications in this field include: Computation into Criticism: A Study of Jane Austen's Novels and an Experiment in Method *(Oxford, Clarendon Press, 1987); "The Reciprocities of Style: Literary Criticism and Literary Statistics,"* Essays and Studies, *n. s. XXXIX (1989), 78—93. He is currently preparing for the Clarendon Press a further book with the working title* Patterns in Rough Ground: a Computer-assisted Study of the Language of Narrative.

Computers and the Humanities **23**: 309—321, 1989.

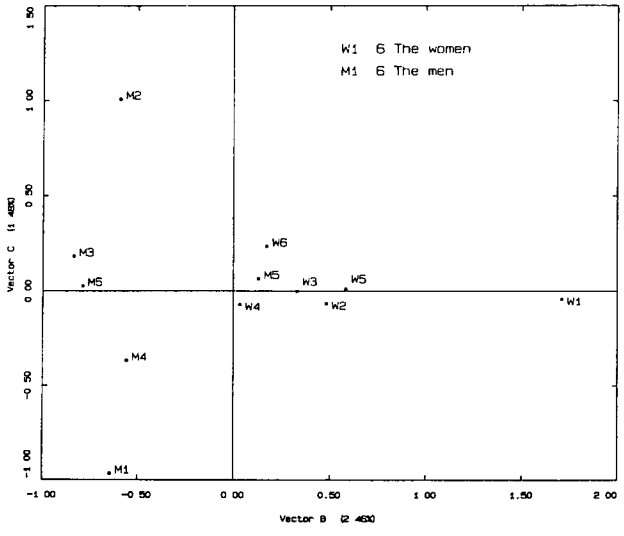
100,000 word-tokens, and the only principle of selection consists in cutting off the word-list at a point where the frequencies begin to thin out. The list of words and the frequency hierarchies for those forty-eight of Jane Austen's characters who speak more then two thousand words apiece are set out in Appendix C, pp. 231—6, of *Computation into Criticism*. Let us examine these graphs briefly before considering the manner in which they are derived.

Graph 1 gives an individual entry for each of the forty-eight major characters. Four of the heroines and three of the heroes lie near the zero-point, with a group of well-spoken neighbours. Most of Jane Austen's comic monsters lie out at the extremities. So do the majority of characters from early novels. Most of the women lie on the eastern side, most men on the western. The more disquisitory characters lie in the south, the more interlocutory in the north. And, transcending but not contradicting this evidence of convention, chronology, and gender, there is a predominance of submissive tempers and colloquial styles in the east, a predominance of dominant tempers and formal styles in the west. A few seeming anomalies, like Mrs. Elton and Mrs. Norris, are given due attention in my book.

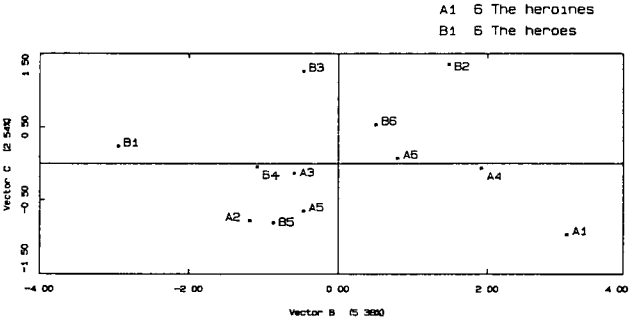
Graph 2 arranges all the characters of Jane Austen's novels in twelve groups — the men and women of each novel. Except for the women of *Northanger Abbey*, the female groups lie close together. The men scatter more widely, with an especially strong contrast between those of *Northanger Abbey* and those of *Sense and Sensibility*. But only the men of *Emma* enter female territory. The remote locations of W1, M1, and M2 signify that more forces than that of gender are at work: but Graph 2 shows, overall, that gender enters into the differentiation of Jane Austen's characters. When the heroes and heroines are singled out, as in Graph 3, gender remains a factor but is less powerful than chronological change.



Graph 1. Jane Austen's major characters.



Graph 2. Gender in Jane Austen.

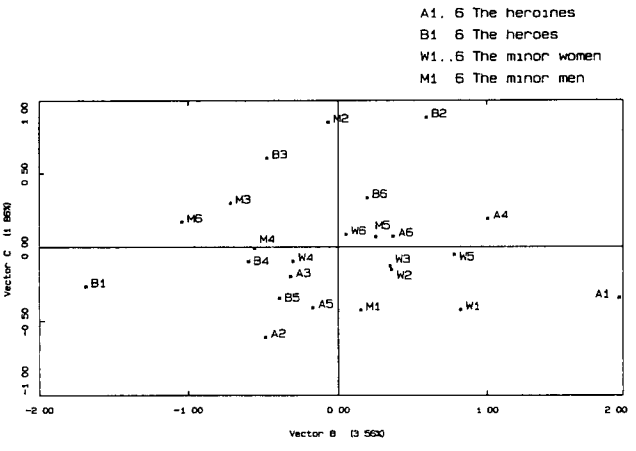


Graph 3. Gender in Jane Austen.

Notice how the gap between the members of each pair grows less and less as Jane Austen's career proceeds. Whatever we choose to make of it, it is clear that a novelist who was capable of distinguishing the idiolects of her leading characters by gender ceased to enforce the distinction in her later work.

In Graph 4, the twenty-four entries comprise all the heroes and heroines, the other men, and the other women. As before, there is evidence of gender-clustering and of some "gender-crossing". But this is outweighed, I suggest, by the overall confusion of a map in which too many different forces are at work to yield a clear picture.

Before turning from the idiolects of literary characters to those of actual writers, I must



Graph 4. Gender in Jane Austen.

describe the methods of analysis that lie behind these graphs.[3] The statistical analysis of literary texts has sometimes been addressed to the presence or absence of rare word-types, sometimes to the relative incidence of very common ones. My method is of the second kind but it differs from others in that, instead of taking a few chosen word-types, it rests on *correlations of the frequency profiles* formed by many common word-types, arranged in the descending order of their frequency in a given text or set of texts. The texts themselves, that is to say, are allowed to dictate which word-types should be studied. Since so few as thirty word-types make up about two-fifths of all the word-tokens most writers employ, so few as fifty almost half, their contribution to the size of any text suggests that they may also contribute to its "shape".

But when there are marked formal differences between the texts to be compared, it is necessary to exclude certain word-types from the analysis. Comparisons of direct and indirect speech are the most striking case in point. The huge differences in the incidence of the personal pronouns and the inflected auxiliary verbs obscure all other resemblances and differences between the texts and, unless they are excluded, the comparisons have little meaning. There was no occasion to exclude these words from the analysis of Jane Austen's dialogue. They are excluded, however, from all the analyses that follow and we shall be treating,

therefore, of a quarter rather than two-fifths of the word-tokens in each text.

Table 1 is a list of "embedded histories", the basis of my later argument. Table 2 shows frequency profiles for ten histories by Henry Fielding, nine by Sarah Fielding, and four by Tobias Smollett. Since we shall treat of the histories as such, the frequencies cited are those of each main narrator, excluding words used by other speakers. All told, we have 87,390 word-tokens from Henry Fielding's histories, 88,536 from Sarah's, and 29,700 from Smollett's. Save for the history of Julian, which makes up most of *A Journey from this World* and from which the first 9468 words are taken as a sample, the whole of each main narrator's text is treated. Save for those of Leonora, Trent, Mrs. Bilson, and Greaves, each history is told by its protagonist. Save for Cleopatra and Octavia, each protagonist addresses other characters. Save for those of Cleopatra, Octavia, and Julian, the histories are all "embedded" in much longer novels.

TABLE 1
List of "histories"

HENRY FIELDING (1707-54)
A Total 87390
A1 Mrs Bennet 14446 (Amelia, 1751)
A2 Capt. Booth 23417 (Amelia, 1751)
A3 Mrs Fitzpatrick 5989 (Tom Jones, 1749)
A4 Mrs Heartfree 5991 (Jonathan Wild, 1743)
A5 Man of the Hill 8320 (Tom Jones, 1749)
A6 Julian 9468 (A Journey, 1743)
A7 Leonora 3839 (Joseph Andrews, 1742)
A8 Miss Mathews 5236 (Amelia, 1751)
A9 Trent 2716 (Amelia, 1751)
A10 Wilson 7968 (Joseph Andrews, 1742)

SARAH FIELDING (1710-68)
B Total 88536
B1 Mrs Bilson 7475 (Countess of Dellwyn, 1759)
B2 Camilla 12271 (David Simple, 1744)
B3 Cleopatra 33460 (Cleopatra & Octavia, 1757)
B4 Cynthia 6460 (David Simple, 1744)
B5 Daniel 1835 (David Simple, 1744)
B6 Dumont 2507 (David Simple, 1744)
B7 Isabelle 13534 (David Simple, 1744)
B8 Octavia 8581 (Cleopatra & Octavia, 1757)
B9 Stainville 2413 (David Simple, 1744)

JANE AUSTEN (1775-1817)
C Total 6305
C1 Col. Brandon 2319 (Sense & Sensibility, 1811)
C2 Willoughby 3986 (Sense & Sensibility, 1811)

JOHN CLELAND (1709-89)
D1 Three whores 6610 (Woman of Pleasure, 1748-9)

WILKIE COLLINS (1824-89)
E Total 35437
E1 Gilmore 8534 (The Woman in White, 1860)
E2 Miss Halcombe 15859 (The Woman in White, 1860)
E3 Hartright 11044 (The Woman in White, 1860)

CHARLES DICKENS (1812-70)
F Total 8518
F1 Convict 3187 (Pickwick Papers, 1837)
F2 Stroller 2146 (Pickwick Papers, 1837)
F3 Miss Wade 3185 (Little Dorrit, 1855-7)

E. L. DOCTOROW (1931-)
G1 Willi 3392 (Lives of the Poets, 1985)

MARIA EDGEWORTH (1767-1849)
H Total 20516
H1 Lady Davenant 8927 (Helen, 1834)
H2 Lady Delacour 11589 (Belinda, 1801)

GEORGE ELIOT (1819-80)
I Total 9812
I1 Mirah 5683 (Daniel Deronda, 1874-6)
I2 Rufus Lyon 4129 (Felix Holt, 1866)

ELIZABETH GASKELL (1810-65)
J Total 10899
J1 Nurse 7797 ("The Old Nurse's Story", 1852)
J2 Peter 3102 (Cranford, 1853)

THOMAS HARDY (1840-1928)
K Total 11058
K1 Lady Baxby 2118 (Noble Dames, 1891)
K2 Lady Icenway 3288 (Noble Dames, 1891)
K3 Lady Petrick 3215 (Noble Dames, 1891)
K4 Selby 2437 (Wessex Tales, 1888)

THOMAS HOLCROFT (1745-1809)
L Total 11906
L1 Miss Wilmot 3367 (Hugh Trevor, 1794)
L2 Wilmot 8539 (Hugh Trevor, 1794)

HENRY JAMES (1843-1916)
M Total 13108
M1 Mrs Bread 4241 (The American, 1877 edn.)
M2 Eustace 8867 ("Master Eustace", 1871)

ERICA JONG (1942-)
N1 Isadora 4082 (Fear of Flying, 1973)

CHARLOTTE LENNOX (1720-1804)
O Total 33181
O1 Cynecia 1378 (Female Quixote, 1752)
O2 Dolly 2412 (Sophia, 1762)
O3 Miss Groves 1744 (Female Quixote, 1752)
O4 Henrietta 18231 (Henrietta, 1758)
O5 Sir George 9416 (Female Quixote, 1752)

ALISON LURIE (1926-)
P1 Zimmern 5485 (Imaginary Friends, 1967)

WALTER SCOTT (1771-1832)
Q Total 10961
Q1 Elspeth 2358 (The Antiquary, 1816)
Q2 Staunton 3546 (Heart of Midlothian, 1818)
Q3 Wandering Willie 5057 (Redgauntlet, 1824)

CHARLOTTE SMITH (1749-1806)
R Total 13752
R1 Adelina 6662 (Emmeline, 1788)
R2 Monimia 7090 (Old Manor House, 1793)

TOBIAS SMOLLETT (1721-71)
S Total 29700
S1 Greaves 10591 (Launcelot Greaves, 1760-2)
S2 Melopoyn 5932 (Roderick Random, 1748)
S3 Mrs Williams 6847 (Roderick Random, 1748)
S4 Zelos 6330 (Ferdinand, Count Fathom, 1753)

TABLE 2
Raw word frequencies for a set of "histories"
(Hierarchy based on the 49 most common word-types in Henry Fielding's "histories")

		Henry Fielding											Sarah Fielding										Smollett				
		Tot H	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	Tot S	S1	S2	S3	S4	S5	S6	S7	S8	S9	Tot T	T1	T2	T3	T4
	the	3995	597	1155	201	299	407	443	217	208	147	321	3494	315	371	1438	212	62	95	534	369	98	1646	728	249	319	350
*	I	2763	468	715	204	179	246	366	12	252	3	318	2468	8	362	1091	230	111	105	258	244	59	682	51	225	225	181
	of	2688	436	715	142	189	294	295	116	144	94	263	2820	224	331	1145	210	61	68	377	330	74	1126	351	210	255	310
	and	2309	344	600	154	157	255	259	111	116	80	233	2697	231	362	1067	154	48	54	414	308	59	1047	387	203	251	206
	a	1834	277	439	140	130	167	223	93	113	62	190	1343	138	202	425	142	28	37	214	120	37	708	276	137	169	126
*	my	1675	361	389	149	89	181	224	1	129	15	137	1777	0	211	840	110	40	82	192	258	44	610	23	172	258	157
	to(inf)	1580	282	429	93	108	141	145	71	119	45	147	2283	165	341	866	164	48	59	370	204	66	512	157	118	125	112
*	was	1493	268	383	97	91	122	196	68	85	70	113	1762	159	262	640	125	47	51	275	171	32	480	164	105	119	92
	in	1491	232	430	96	90	134	155	71	90	54	139	1609	122	226	634	125	33	46	227	146	50	570	202	127	122	119
	to(prep)	1338	230	335	97	83	136	151	57	78	38	135	1238	118	151	490	80	27	38	179	120	35	452	136	104	117	95
*	he	1194	263	237	114	122	127	88	67	59	78	39	1167	80	126	408	61	6	45	288	99	54	476	251	110	78	37
*	had	1065	210	283	55	77	108	109	45	60	34	84	1123	117	155	395	82	39	26	177	94	38	279	95	56	64	64
*	me	1037	198	219	76	107	82	168	0	69	0	118	933	1	145	358	84	38	56	132	94	25	265	4	100	123	38
	with	976	148	233	54	83	83	117	54	64	30	110	937	98	111	360	66	18	20	148	95	21	428	151	90	101	86
	which(rp)	968	164	268	63	79	93	101	49	33	32	86	515	57	62	206	33	12	10	71	49	15	230	54	57	52	67
*	his	884	214	154	45	80	107	65	71	45	72	31	1334	96	108	586	35	2	27	282	167	31	506	264	99	77	66
*	her	838	141	354	26	8	11	56	92	31	22	97	992	281	196	136	47	7	10	274	38	3	153	68	3	26	56
	as	825	113	204	48	54	117	105	47	30	27	80	997	104	145	397	81	23	24	113	87	23	226	88	47	50	41
	it	773	130	211	67	40	68	83	30	64	15	65	795	59	141	241	81	13	31	128	78	23	134	39	48	23	24
	that(cj)	703	131	229	39	53	56	67	23	41	19	45	848	57	90	388	37	16	25	136	70	29	251	71	75	59	46
	at	693	92	200	46	44	80	71	29	47	24	60	539	45	73	208	41	9	20	97	31	15	204	83	55	35	31
	not	660	116	174	50	40	45	73	22	53	19	68	633	74	101	194	54	14	20	96	65	15	206	69	53	42	42
	this	644	110	147	61	44	54	73	15	45	27	68	510	31	78	193	25	9	21	77	54	22	216	59	60	43	54
*	she	601	141	231	20	2	12	32	64	24	19	56	657	157	209	58	50	3	7	161	12	0	86	24	1	25	36
	but	547	98	114	47	56	54	69	25	31	9	44	706	75	102	212	63	16	25	140	55	18	172	71	33	42	26
	for(prep)	516	86	129	36	33	38	60	28	40	17	49	647	58	102	211	41	9	23	123	60	20	199	60	34	56	49
	on(prep)	467	88	118	27	46	34	56	20	31	15	32	492	32	56	210	30	12	11	90	40	11	122	57	14	35	16
*	you	459	72	137	74	11	31	13	7	68	0	46	104	0	27	3	14	1	31	11	1	16	69	25	19	8	17
	all	452	78	123	31	30	42	38	25	25	19	41	563	43	84	217	29	25	17	99	33	16	127	54	16	30	27
	by(prep)	448	65	107	27	27	53	65	21	11	22	50	566	59	62	253	29	14	14	64	61	10	183	61	30	39	53
*	him	446	94	84	32	43	65	44	20	26	22	16	988	43	97	472	32	3	14	206	84	37	179	94	30	42	13
	from	439	87	128	32	36	37	40	18	18	12	31	466	54	56	174	16	9	9	65	71	12	120	37	19	29	35
*	have	435	65	104	52	17	35	49	15	43	7	48	446	16	81	145	62	6	17	59	41	19	88	27	15	18	28
	very	396	59	83	41	46	41	47	15	21	21	22	219	14	37	77	26	4	5	38	14	4	60	16	25	11	8
	be	393	59	108	37	33	38	32	12	33	12	29	544	29	96	204	62	7	9	73	52	12	119	44	28	27	20
	for(cj)	352	69	87	35	15	44	29	5	18	17	33	360	16	58	138	46	9	8	52	28	5	42	18	9	9	6
*	we	341	58	120	16	59	30	30	1	2	5	20	246	1	103	31	15	5	4	47	18	22	35	14	1	12	8
	so(av d)	333	51	94	34	22	32	35	18	15	8	24	452	55	55	154	43	7	12	76	33	17	93	36	19	22	16
	an	326	47	81	20	20	33	41	21	23	15	25	266	22	30	96	24	5	12	37	33	7	127	39	18	38	32
	now	324	48	92	20	20	43	37	9	26	8	21	147	4	25	51	9	3	4	29	18	4	34	15	8	5	6
	that(dem)	319	55	86	25	21	25	34	15	29	10	19	273	23	43	95	20	10	7	47	21	7	104	26	23	18	37
	who(rp)	301	36	70	14	18	47	49	21	0	11	35	212	33	0	104	0	10	7	21	27	10	158	50	43	29	36
*	is	297	44	84	28	14	21	35	12	24	4	31	161	3	31	30	44	2	5	18	17	11	51	22	6	8	15
	no	290	49	69	32	18	29	23	10	20	11	29	235	21	42	77	22	7	4	28	25	9	68	23	10	18	17
	when	290	49	86	22	27	20	27	12	21	8	18	276	21	35	112	16	8	13	43	22	6	90	25	24	24	17
*	could	274	54	78	22	17	26	23	6	16	4	28	509	41	84	165	38	11	19	81	57	13	87	33	20	22	12
	more	262	40	81	21	12	22	22	11	13	7	33	208	28	23	74	26	1	6	28	18	4	49	14	10	22	3
*	were	260	46	74	12	22	23	32	12	8	6	25	326	35	60	128	19	7	6	36	31	4	84	48	8	9	19
	than	259	29	80	15	23	21	23	13	11	9	35	205	25	32	71	23	3	3	33	10	5	45	15	9	14	7

Table 3 makes the profiles more intelligible by translating the raw figures into percentages of all the word-tokens in each text. Table 4 is a correlation-matrix (based on Pearson's product-moment formula) for the nineteen Fielding histories. In Table 4 (where such values as +0.995 and +0.970 are more concisely represented as 995 and 970) the coefficients all run very high. That is because the shape of the language makes it impossible, in any ordinary sort of English, for word-types like "the" and "of" to slip towards the bottom of our list and be replaced at the top by word-types like "than" and "more". Yet, though they run so high, the coefficients of Table 4 fall into clear patterns. In the top left, where Henry Fielding's histories are correlated, the coefficients are especially high. So, too, in the bottom right, for the correlations among Sarah Fielding's histories. But, in the bottom left, where Sarah's are correlated with Henry's, they are distinctly lower.

When a correlation matrix is subjected, in turn, to what is known as "eigen analysis", a form of multilinear regression, the overall pattern of interrelationships can be clearly displayed in graphs like those already offered. The function of eigen analysis is twofold. It translates the manifold patterns of a matrix into a succession of "eigen vectors" and it assigns an exact weighting, an "eigen value", to each vector. The first vector arrays the coefficients in their most mutually consistent sequence, a "line of best fit" like the longest axis of an egg-shaped field. The second vector arrays them in the most consistent residual sequence, like the longest transverse axis of the field. And so on. The successive eigen values show whether the egg aspires to the shape of a cigar, that

TABLE 3
Incidences (as text-percentages) of main data in Table 2

	Henry Fielding											Sarah Fielding									
	Tot H	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	Tot S	S1	S2	S3	S4	S5	S6	S7	S8	S9
the	4.57	4.13	4.93	3.36	4.99	4.89	4.68	5.65	3.97	5.41	4.03	3.95	4.21	3.02	4.30	3 28	3.38	3.79	3.95	4.30	4.06
* I	3.16	3.24	3.05	3.41	2.99	2.96	3.87	0.31	4.81	0.11	3.99	2.79	0.11	2.95	3.26	3 56	6.05	4.19	1.91	2.84	2 45
of	3.08	3.02	3.05	2.37	3.15	3.53	3.12	3.02	2.75	3.46	3.30	3.19	3.00	2.70	3.42	3 25	3.32	2.71	2.79	3.85	3.07
and	2.64	2.38	2.56	2.57	2.62	3.06	2.74	2.89	2.22	2.95	2.92	3.05	3.09	2.95	3.19	2.38	2.62	2.15	3.06	3.59	2.45
a	2.10	1.92	1.87	2.34	2.17	2.01	2.36	2.42	2.16	2.28	2.38	1.52	1.85	1.65	1.27	2.20	1.53	1.48	1.58	1.40	1.53
* my	1.92	2.50	1.66	2.49	1.49	2.18	2.37	0.03	2.46	0.55	1.72	2.01	0.00	1.72	2.51	1.70	2.18	3.27	1.42	3.01	1.82
to(inf)	1.81	1.95	1.83	1.55	1.80	1.69	1.53	1.85	2.27	1.66	1.84	2.58	2.21	2.78	2.59	2.54	2.62	2.35	2 73	2.38	2.74
* was	1.71	1.86	1.64	1.62	1.52	1.47	2.07	1.77	1.62	2.58	1.42	1.99	2.13	2.14	1.91	1.93	2 56	2.03	2 03	1.99	1.33
in	1.71	1.61	1.84	1.60	1.50	1.61	1.64	1.85	1.72	1.99	1.74	1.82	1 63	1.84	1.89	1.93	1.80	1 83	1.68	1.70	2 07
to(prep)	1.53	1.59	1.43	1.62	1.39	1.63	1.59	1.48	1.49	1.40	1.67	1.40	1.58	1 23	1.46	1.24	1.47	1.52	1 32	1.40	1.45
* he	1.37	1.82	1.01	1.90	2.04	1.53	0.93	1.75	1.13	2.87	0.49	1.32	1.07	1 03	1 22	0 94	0.33	1 79	2 13	1.15	2.24
* had	1.22	1.45	1.21	0.92	1.29	1.30	1.15	1.17	1.15	1.25	1.05	1.27	1.57	1 26	1.18	1 27	2.13	1.04	1 31	1.10	1 57
* me	1.19	1.37	0.94	1.27	1.79	0.99	1.77	0.00	1.32	0.00	1.48	1.05	0.01	1.18	1.07	1.30	2.07	2.23	0.98	1.10	1.04
with	1.12	1.02	1.00	0.90	1.39	1.00	1.24	1.41	1.22	1.10	1.38	1.06	1.31	0 90	1.08	1.02	0.98	0 80	1.09	1.11	0 87
which(rp)	1.11	1.14	1.14	1.05	1.32	1.12	1.07	1.28	0.63	1.18	1.08	0.58	0.76	0.51	0.62	0.51	0.65	0.40	0.52	0 57	0 62
* his	1.01	1.48	0.66	0.75	1.34	1.29	0.69	1.85	0.86	2.65	0.39	1.51	1.28	0.88	1.75	0.54	0.11	1.08	2.08	1.95	1 28
* her	0.96	0.98	1.51	0.43	0.13	0.13	0.59	2.40	0.59	0.81	1.22	1.12	3.76	1.60	0.41	0.73	0.38	0.40	2.02	0.44	0.12
as	0.94	0.78	0.87	0.80	0.90	1.41	1.11	1.22	0.57	0.99	1.00	1.13	1.39	1.18	1.19	1.25	1.25	0 96	0 83	1.01	0 95
it	0.88	0.90	0.90	1.12	0.67	0.82	0.88	0.78	1.22	0.55	0.82	0.90	0.79	1.15	0.72	1 25	0.71	1 24	0.95	0 91	0 95
that(cj)	0.80	0.91	0.98	0.65	0.88	0.67	0.71	0.60	0.78	0.70	0.56	0.96	0.76	0.73	1.16	0 57	0.87	1.00	1.00	0.82	1 20
at	0.79	0.64	0.85	0.77	0.73	0.96	0.75	0.76	0 90	0.88	0.75	0.61	0.60	0.59	0.62	0 63	0.49	0 80	0.72	0.36	0 62
not	0.76	0.80	0.74	0 83	0.67	0.54	0.77	0.57	1 01	0.70	0.85	0.71	0.99	0 82	0.58	0.84	0.76	0 80	0.71	0.76	0 62
this	0.74	0.76	0 63	1.02	0.73	0 65	0.77	0.39	0.86	0.99	0.85	0.58	0.41	0 64	0.58	0.39	0.49	0 84	0.57	0 63	0 91
* she	0.69	0.98	0 99	0.33	0.03	0.14	0.34	1.67	0.46	0.70	0.70	0.74	2.10	1 70	0 17	0 77	0.16	0 28	1 19	0 14	0.00
but	0.63	0 68	0.49	0.78	0.93	0.65	0.73	0.65	0.59	0.33	0.55	0.80	1 00	0 83	0 63	0.98	0 87	1.00	1 03	0 64	0 75
for(prep)	0.59	0 60	0.55	0.60	0.55	0.46	0.63	0.73	0.76	0 63	0.61	0.73	0.78	0.83	0 63	0.63	0.49	0.92	0 91	0 70	0.83
on(prep)	0.53	0 61	0.50	0.45	0.77	0.41	0.59	0.52	0.59	0 55	0.40	0.56	0 43	0.46	0 63	0.46	0 65	0.44	0 66	0 47	0.46
* you	0.53	0.50	0.59	1.24	0.18	0.37	0.14	0.18	1.30	0 00	0.58	0.12	0 00	0.22	0 01	0.22	0 05	1.24	0 08	0 01	0 66
all	0.52	0.54	0 53	0.52	0.50	0.50	0.40	0.65	0.48	0.70	0.51	0.64	0 58	0.68	0 65	0.45	1 36	0.68	0 73	0.38	0 66
by(prep)	0 51	0.45	0 46	0.45	0.45	0.64	0.69	0.55	0.21	0.81	0.63	0.64	0 79	0.51	0 76	0 45	0 76	0 56	0.47	0 71	0.41
* him	0.51	0.65	0 36	0.53	0.72	0.78	0.46	0.52	0.50	0.81	0.20	1.12	0.58	0 79	1.41	0 50	0.16	0.56	1.52	0.98	1 53
from	0.50	0.60	0 55	0.53	0.60	0 44	0.42	0 47	0.34	0.44	0.39	0.53	0.72	0.46	0.52	0 25	0.49	0.36	0.48	0.83	0 50
* have	0.50	0 45	0.44	0.87	0.28	0.42	0.52	0.39	0.82	0.26	0.60	0.50	0.21	0.66	0.43	0.96	0.33	0.68	0.44	0.48	0 79
very	0.45	0 41	0.35	0.68	0.77	0.49	0.50	0.39	0.40	0.77	0.28	0.25	0.19	0.30	0.23	0.40	0.22	0 20	0 28	0.16	0 17
be	0.45	0 41	0.46	0.62	0.55	0.46	0.34	0.31	0.63	0.44	0.36	0.61	0 39	0.78	0 61	0.96	0.38	0 36	0.54	0 61	0 50
for(cj)	0.40	0.48	0.37	0 58	0.25	0.53	0.31	0.13	0.34	0.63	0.41	0.41	0 21	0.47	0 41	0 71	0 49	0 32	0.38	0.33	0 21
* we	0.39	0.40	0 51	0 27	0.98	0.36	0.32	0.03	0.04	0.18	0.25	0.28	0.01	0.84	0.09	0.23	0.27	0 16	0.35	0.21	0.91
so(av d)	0.38	0.35	0.40	0.57	0.37	0.38	0.37	0.47	0.29	0.29	0.30	0.51	0.74	0.45	0.46	0.67	0.38	0 48	0.56	0.38	0.70
an	0.37	0.33	0.35	0.33	0.33	0.40	0.43	0.55	0.44	0.55	0.31	0.30	0.29	0.24	0.29	0.37	0.27	0 48	0 27	0.38	0 29
now	0.37	0.33	0.39	0.33	0.33	0.52	0.39	0.23	0.50	0.29	0.26	0.17	0.05	0 20	0.15	0.14	0.16	0.16	0.21	0.21	0 17
that(den)	0.37	0.38	0.37	0.42	0.35	0.30	0.36	0.39	0.55	0.37	0.24	0.31	0.31	0 35	0.28	0.31	0.54	0.28	0.35	0.24	0 29
who(rp)	0.34	0.25	0.30	0.23	0.30	0.56	0.52	0.55	0.00	0.41	0.44	0.24	0.44	0.00	0.31	0.00	0.54	0.28	0.16	0 31	0 41
* is	0.34	0.30	0 36	0.47	0.23	0.25	0.37	0.31	0.46	0.15	0.39	0.18	0.04	0.25	0.09	0 68	0.11	0.20	0.13	0.20	0.46
no	0.33	0.34	0.29	0.53	0.30	0.35	0.24	0.26	0.38	0.41	0.36	0 27	0 28	0.34	0.23	0.34	0.38	0.16	0.21	0 29	0.37
when	0.33	0.34	0.37	0.37	0.45	0.24	0.29	0.31	0.40	0.29	0.23	0.31	0.28	0.29	0.33	0 25	0.44	0 52	0.32	0 26	0.25
* could	0.31	0.37	0.33	0 37	0.28	0.31	0.24	0.16	0.31	0.15	0.35	0.57	0.55	0 68	0.49	0 59	0.60	0 76	0 60	0.66	0.54
more	0.30	0 28	0.35	0.35	0.20	0.26	0.23	0.29	0.25	0.26	0.41	0.23	0.37	0.19	0.22	0 40	0.05	0 24	0.21	0.21	0 17
* were	0.30	0.32	0.32	0.20	0.37	0 28	0.34	0.31	0 15	0.22	0.31	0.37	0.47	0.49	0.38	0 29	0.38	0 24	0.27	0.36	0 17
than	0.30	0 20	0.34	0.25	0.38	0.25	0.24	0.34	0.21	0.33	0.44	0.23	0.33	0 26	0.21	0 36	0.16	0.12	0.24	0.12	0 21

TABLE 4
Pearson correlation matrix for main data in Table 2
(based on the 33 unasterisked word types of Table 1)

	H																			
H1	995	H1																		
H2	995	989	H2																	
H3	970	968	951	H3																
H4	989	983	985	949	H4															
H5	989	977	982	952	976	H5														
H6	994	984	984	967	986	988	H6													
H7	986	971	985	942	984	977	987	H7												
H8	969	974	963	958	951	936	954	941	H8											
H9	987	974	985	949	980	983	985	980	939	H9										
H10	985	980	968	973	964	977	984	964	957	969	H10									
S	962	967	955	931	940	954	944	935	950	934	958	S								
S1	969	966	959	939	953	964	965	961	935	941	969	982	S1							
S2	912	924	898	917	880	898	890	875	933	873	923	979	946	S2						
S3	959	963	958	913	940	957	941	934	934	939	949	995	974	959	S3					
S4	927	934	910	927	899	913	911	892	944	890	937	959	938	968	936	S4				
S5	926	939	915	890	901	922	908	897	910	900	930	973	951	954	970	936	S5			
S6	951	959	949	922	924	930	934	925	961	918	933	973	954	950	964	941	947	S6		
S7	951	959	946	927	935	934	932	929	955	918	941	992	972	976	982	947	958	973	S7	
S8	956	960	949	922	936	957	944	926	930	935	956	988	974	954	987	934	955	946	970	S8
S9	955	965	955	921	932	935	933	928	955	925	940	983	959	957	979	943	959	984	978	960

of an oval plaque, or (when no vector predominates) that of a sphere.

In cases like those before us, the uniformity of the coefficients ensures that the first vector is predominant — and "uninteresting". For, in such cases, Vector A indicates that all the profiles are much alike and attracts 90% or more of the eigen values to its expression of that truth. It thus reminds us of what any of us (lacking the monstrous tolerance of a computer) would normally take for granted: that texts written in a highly conventional literary form by writers who are close contemporaries are likely to resemble each other closely. In such texts, especially, it is quite improbable that the most common words will slide down the frequency list and, supplanted by their "lesser" rivals, disrupt the pattern of correlations.

When the evidence of Vector A has been registered and set aside, the lesser vectors display (and attach eigen values to) any residual patterning of the matrix, any tendency towards either a distinctive clustering or an indiscriminate scattering. The patterns yielded by Vectors B, C, and D are often of great interest. (After that, the force of the evidence quickly falls away: Vector E and its successors attract too little eigen value to repay close examination.) To bring such patterns into a proportionate relationship with each other and to make properly scaled graphs, it is necessary, finally, to multiply every entry in each vector by the eigen value of that vector. Table 5 is derived, in this manner, from Table 4 and is the basis of Graph 5.

Such graphs make vivid pictures. Pictures of what? Of relationships among coefficients — which are yielded by correlating the original frequency profiles — which reflect habits of usage in the most common word-types of each text — which should, I contend, betoken characteristic features of the texts themselves. The value of that contention is tested by the graphs.

Graph 5 introduces my second set of specimens. The horizontal separation into two authorial clusters, yielded by Vector B, is its salient feature. The vertical axis represents Vector C, which offers a new determinant. At the upper edge are the most "Bakhtinian" texts, those whose speakers (Mrs. Fitzpatrick, Miss Mathews, and Cynthia) most enter into "dialogic relationships" with other characters. At the lower edge are three monologues: Cleopatra, Octavia, and Daniel. None of Henry Fielding's texts lie near this monologic edge. Graph 5 also suggests that, as their careers continued, the Fieldings came to write somewhat — only somewhat — more like each other. Those of Sarah's texts that lie least far from Henry's are all late work. Among his later pieces, leaning towards hers, that of Wilson (H10) is the only chronological oddity: but the location of Wilson's history is altered by its sententious closing pages, which can be shown to carry it away from a more "natural" location. The evidence of chronological change in the Fieldings, however, is less firmly established, at this stage of my research, than its equivalent in Jane Austen.[4]

The calibrations marked along the sides of the graphs stand in a linear relation to each other and allow exact comparisons between the vectors. In Graph 5, the authorial difference has more than twice the weight of the difference between monologic and dialogic texts. It is also possible to make rough comparisons between different graphs. These cannot be exact because different sets of entries affect each other's relative locations in a complex system of mutual attraction and repulsion.

Graph 6 is derived from the same nineteen texts as Graph 5 but their "sanctity" as texts is violated by breaking each of them into successive segments of two thousand words, rounding any residue to the nearest two thousand, establishing a frequency profile for each segment, and proceeding as before. The graph contains eighty-eight entries, forty-four apiece from Henry and Sarah Fielding. Although a few entries, including the closing phase of Wilson's history, lie near the boundary between the two authorial clusters, *not one* of the eighty-eight leaps the border and invades the territory of the other author. Statistical analysis never enables the mind to repose on what Johnson called the stability of truth: but the degree of probability vested in patterns like those of Graph 6 is formidable indeed.

Graph 7, which is derived from the appropriate columns of Table 2, shows an equally clear authorial distinction between the histories of

TABLE 5
Eigen Vectors derived from Table 4.
(First 4 only)

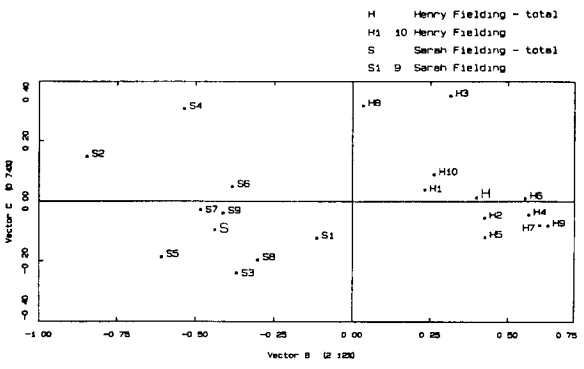
A	B	C	D
-0.221683	0.188395	0.017091	-0.033919
-0.221713	0.109719	0.052414	-0.107274
-0.220092	0.200924	-0.076351	-0.225364
-0.215785	0.148549	0.475062	0.235303
-0.217981	0.267786	-0.061832	-0.096818
-0.219233	0.201249	-0.165965	0.204367
-0.219363	0.262168	0.012652	0.090881
-0.217234	0.284140	-0.112183	-0.079832
-0.217543	0.015979	0.431003	-0.340466
-0.217083	0.296708	-0.113292	0.013310
-0.219807	0.123830	0.119774	0.331983
-0.220904	-0.206709	-0.132744	0.035058
-0.219936	-0.053463	-0.169992	0.200704
-0.213302	-0.399588	0.199202	0.173196
-0.219552	-0.173818	-0.323834	-0.009403
-0.213527	-0.253056	0.416600	0.249948
-0.214154	-0.286979	-0.253428	0.132092
-0.217409	-0.180887	0.065199	-0.470133
-0.218891	-0.228194	-0.039180	-0.151252
-0.218534	-0.142411	-0.265981	0.231948
-0.218583	-0.194400	-0.054857	-0.374013

Eigen Values for the first 4 vectors.
(expressed as percentages)

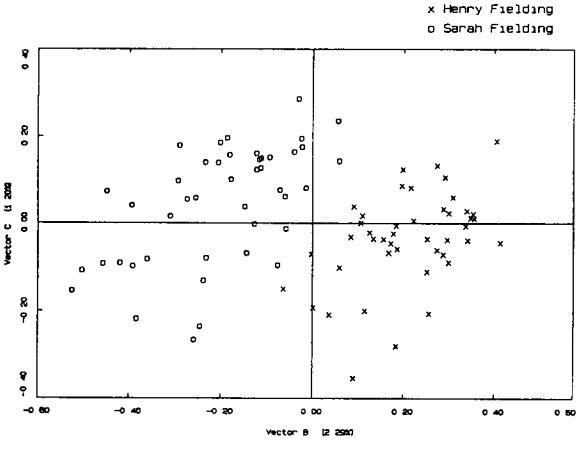
A	B	C	D
95.3020	2.1228	0.7420	0.5079

Product of each entry by the relevant eigen value.

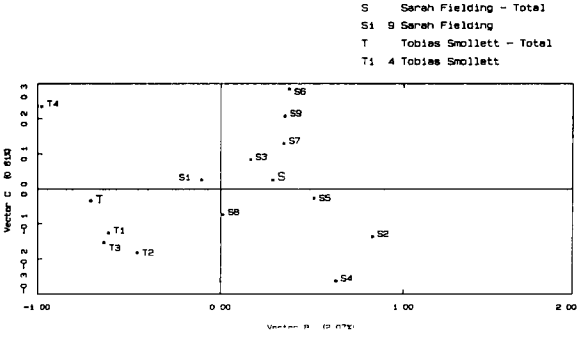
A	B	C	D	Key
-21.1264	0.399398	0.012648	-0.017298	H
-21.1292	0.232604	0.038786	-0.054710	H1
-20.9748	0.425960	-0.056500	-0.114936	H2
-20.5643	0.314924	0.351546	0.120005	H3
-20.7736	0.567705	-0.045755	-0.049377	H4
-20.8929	0.426649	-0.122814	0.104227	H5
-20.9053	0.555795	0.009363	0.046349	H6
-20.7024	0.602377	-0.083015	-0.040714	H7
-20.7319	0.033875	0.318942	-0.173637	H8
-20.6880	0.629021	-0.083836	0.006788	H9
-20.9476	0.262519	0.088632	0.169311	H10
-21.0522	-0.438222	-0.098230	0.017879	S
-20.9599	-0.113342	-0.125794	0.102359	S1
-20.3277	-0.847125	0.147410	0.088330	S2
-20.9233	-0.368494	-0.239637	-0.004795	S3
-20.3492	-0.536478	0.308284	0.127473	S4
-20.4089	-0.608396	-0.187537	0.067367	S5
-20.7191	-0.383481	0.048247	-0.239768	S6
-20.8603	-0.483772	-0.028993	-0.077138	S7
-20.8263	-0.301911	-0.196826	0.118293	S8
-20.8310	-0.412128	-0.040594	-0.190747	S9



Graph 5. "Histories" by Henry and Sarah Fielding.

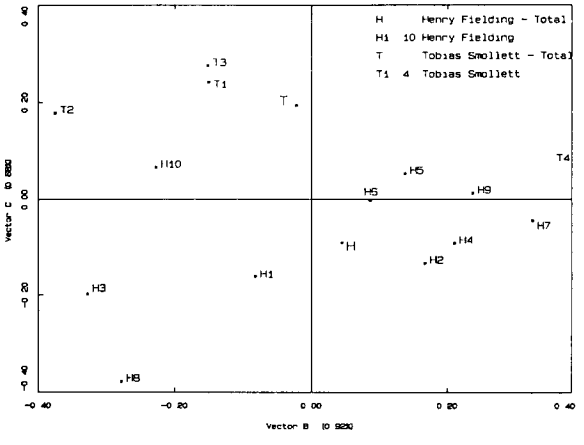


Graph 6. "Histories" by Henry and Sarah Fielding.

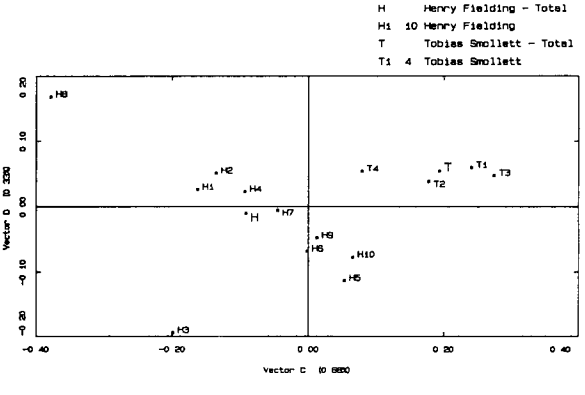


Graph 7. "Histories" by Sarah Fielding and Smollett.

Sarah Fielding and those of Tobias Smollett, a contemporary of the Fieldings. But Graphs 8 and 9 give a fresh turn to the evidence. Both the overall shape of Graph 8 and the location of its entries show that these two authors — Henry Fielding and Smollett — are less easily distinguishable than either of the other pairs. T4, the history of Zelos in *Ferdinand, Count Fathom*, lies on the very edge of Henry Fielding's territory and none of the other Smollett entries lies far away. Since the authorial determinant now appears on Vector C, the vertical axis of the graph, it carries less eigen value, less status as a determinant, than in the other graphs.



Graph 8. "Histories" by Fielding and Smollett.

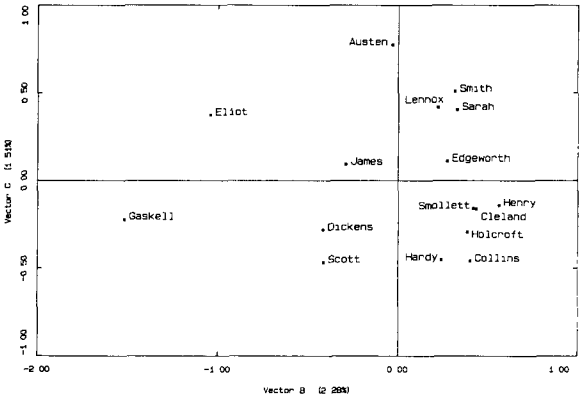


Graph 9. "Histories" by Fielding and Smollett.

This evidence of a lesser difference between Henry Fielding and Smollett than between the other pairs is confirmed by Graph 9. The two authorial clusters now stand apart but we have been obliged to turn to Vectors C and D (with eigen values of only 0.68% and 0.33%) to make them do so. The implication is that "authorial signatures" can be identified even when two writers have many linguistic habits in common but, on the other hand, that the differences between signatures are not always primary points of difference. (When the texts of all three writers are broken into two thousand word segments in the manner of Graph 6, the clusters for Sarah Fielding and Smollett stand well apart: but those for Smollett and Henry Fielding interpenetrate each other to a small extent even when Vectors C and D are plotted.)

Consider the larger implications of this second set of graphs. All of the texts are couched in one literary form, the embedded (or virtually embedded) history. They were all written within a period of twenty years. Nothing in the patterns shows any sign of an Austenish ability to differentiate characters by gender or by temperament. All three sets bear distinct authorial signatures. Two of the sets are by a brother and sister who worked together for much of their lives and sometimes contributed to each other's writings. But those of the English brother and sister resemble each other less closely than the brother's resemble those of his Scottish rival. The question is whether it is fanciful to postulate that the individual differences among the three may express, in part, the difference of gender and and that this may override other differences like that of nationality.

Graph 10 gives clear support to the hypothesis. When more than fifty texts by sixteen novelists of the eighteenth and nineteenth centuries are taken in authorial groups, the results amount to a clustering by chronology and by gender. On the vertical axis of the graph, most of the authors group themselves by gender. The fact that both of James's texts are narrated by women may account for his apparent "feminization": that possibility will be tested by adding male narrators to his set. Maria Edgeworth's unusual education may account for her degree of "masculinization" (which can also be seen when her own letters are compared with those of her contemporaries). In well populated graphs of this kind, the principal vectors often begin to work hand in hand and yield a diagonally oriented separation of the clusters. Mrs. Gaskell, that is to say, lies at the extremity of a diagonal clustering of the women writers and is not (as the vertical axis alone would suggest) to be seen as "masculinized".



Graph 10. "Histories" by Sixteen Novelists.

On a line running from the north-east to the south-west of Graph 10, each cluster arrays itself in rough chronological order, with stronger evidence of change among the women. The most anomalous locations from a chronological point of view, those of Walter Scott and Wilkie Collins, are probably due to my present choice of texts and reflect the capacity of these authors to distinguish between formal and colloquial speakers. That possibility will be tested by adding Gabriel Betteredge, of *The Moonstone*, to Gilmore, Hartright, and Marian Halcombe. Scott's set of three texts yokes the formal English narrative of George Staunton uneasily with the Scottish colloquialism of Elspeth and Wandering Willie; and, when they are treated separately, the first lies far from the other two. Although James's chronological location is not obviously anomalous, it, too, is arrived at by combining the frequency profiles of very different narratives. The only other notable anomaly arises with Mrs. Lennox's Dolly. When her five texts are treated separately, four lie in a closely knit group but that of Dolly, a simple servant-girl, lies far out towards the western extremity of the map. This may be the effect either of an uncommonly suc-

cessful exercise in the differentiation of character or of a different authorial signature. Although *The History of Harriot and Sophia*, from which Dolly's history is drawn, is "nominally by [Mrs. Lennox, it] contains many contributions by her friends".[5]

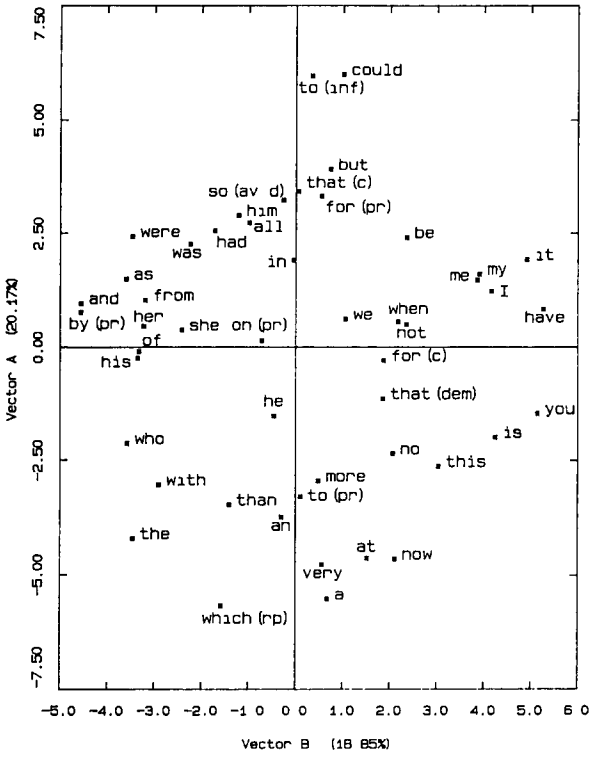
More interesting than the anomalies, however, is the overall coherence of the graph. Its chronological tendency towards the west shows a progressive relaxation of the disciplines of Augustan prose. The fact that this tendency is more marked among the women writers may be taken as evidence of their freer and more active role either in lowering the erstwhile standards of the language or in developing a greater "naturalness" in this form of writing. Leaving all such phallocentric and gynocratic judgments to those who value them, one can sketch a more verifiable line for further inquiry. If it emerged that less educated male writers tended to join the women, that would suggest that the present group of male writers were influenced by their more formal education: but if, in spite of their lack of formal education, those others remained among the men, that would point to broader forms of cultural difference between the genders. It is true, at least, that my first studies of twentieth-century writers alter the whole picture. A small specimen from Erica Jong finds its place in the western territory between George Eliot and Mrs. Gaskell. Two specimens from Alison Lurie and E. L. Doctorow lie side by side, between Hardy and Scott. A specimen of Georgette Heyer's Regency pastiche lies isolated beyond the extreme north-western corner of Graph 10. These specimens are too few and slight to warrant inclusion in the graph and, Heyer aside, their more inward looking forms of "retrospective personal narrative" are not quite of the same genre as the old "embedded history": yet, if they are typical of our time, they suggest that the breakdown of gender-divisions in education over the last hundred years may allow twentieth-century writers less gender-defined forms of expression than were available to their predecessors.

This result is in keeping with the main conclusions emerging from seven years of work on texts of different kinds. The most powerful differentia I have encountered is the simple contrast between direct and indirect speech. When that is brought under control, either by an appropriate choice of texts or by the exclusion of the words that it most affects, differences of genre come into their own. When they too are controlled, chronological differences, authorial differences, and differences of authorial "class" distinguish text from text. Some of the most admired novelists in the canon, Jane Austen notable among them, are able to distinguish appropriately among the idiolects of their characters — whether as individual "personalities" or as members of particular classes. My recent work on published letters, finally, shows very little difference between the letters included in novels and the personal letters of those novelists themselves.

The gist of it all lies in the phrase I have taken from Andrew Marvell for my title. In the Ocean of the language, the deep currents of the common words enable each Kind to find its own Resemblance — but only when suitable mates are made available. But, transcending the measurable determinants of literary style, are the Other Worlds and Other Seas of human individuality. This claim is not a stubbornly romantic defence of the ineffable but a sober assessment of the limitations of such evidence as mine. Though their "authorial signatures" can be distinguished and though they are imbued with appropriate marks of "class", the great novels of the last two centuries bear more subtle and idiosyncratic messages than those. Such messages, of course, are the province of the literary critic, the "stylistician" and (first and last) the responsive reader.

By way of a coda, nevertheless, it is possible to discover which words carry most weight in any given comparison of texts and thus to extract plainer evidence than that of a set of frequency profiles. If, instead of correlating the *columns* of (say) Table 2, we translate them into the percentage-values of Table 3 and then correlate the *rows*, the matrix of coefficients will show which word-types act in consonance with each other and which do not. It is now unnecessary to exclude any word-types from the set: each can be allowed to locate itself with those that behave like it and to stand apart from those that do not. An eigen matrix will then establish a set of principal vectors, as before, and these can then be graphed.

Graph 11 is a characteristic product of this altered approach to the analysis. Vector A no



Graph 11. Henry & Sarah Fielding: Word Plot.

longer dominates the eigen values because the interrelationships among the word-types are more complex than the interrelationships among an essentially similar set of frequency profiles. In Graph 11, Vector A arrays the word-types of Table 3 on authorial lines. Those which (as Table 3 shows) consistently run higher in Henry Fielding's "histories" than in Sarah's lie towards the bottom of the graph: those more characteristic of Sarah lie towards the top. Vector B, meanwhile, distinguishes between the more personally and the more impersonally couched texts: the personal pronouns and the auxiliary verbs are the most obvious markers of this second kind of contrast.

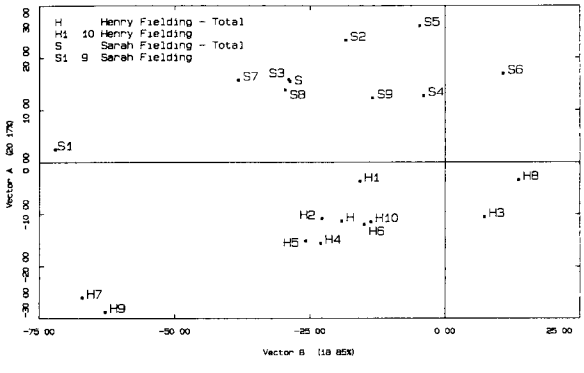
This is no place to trace out the stylistic contrasts to which Graph 11 draws attention. Suffice it to say, by way of illustration, that Henry Fielding's predilection for "which" points to a different kind of syntax from that signified by Sarah's predilection for "but" and the conjunction "that"; that his "very" points to a plainer form of emphasis than the long clauses of result marked by her "so . . . that . . . "; that her disposition towards "could" and the infinitive form of "to" betokens a use of complex tenses; and that his use of "the", "a", and "an" points to a greater emphasis than hers on *things*. If a different pair of writers were compared, the patterning of such a graph would yield quite another set of contrasts and pose other questions for the analyst of style.

If the eigen matrix on which Graph 11 is based is multiplied by the percentage matrix of Table 3, the set of texts is brought back into view. Graph 12 shows the result of this last process. The vertical array shows the now familiar separation of the Fieldings' nineteen "histories" into distinct authorial clusters. The horizontal array, meanwhile, distinguishes most sharply between the three texts in which a third-person narrator holds the stage and the three in which the protagonist is most preoccupied with self.

It is possible, finally, to "massage" the data by excluding those word-types which exert no active influence on the clustering. To use evidence like that of Graph 11 as the basis for making an informed selection of the most relevant data is a legitimate process in statistical analysis. But, for my present purpose and for readers who may suspect a circularity, it is better to avoid any suggestion of tampering with results that speak plainly enough as they stand.

Notes

[1] The research that lies behind this paper could not have been conducted without a great deal of support. The project



Graph 12. Henry & Sarah Fielding: Plot of Transposed Matrix

has been generously funded by the Australian Research Committee and the University of Newcastle, N. S. W. It has made use of the Oxford Concordance Package (OCP) and MINITAB (University of Pennsylvania) and has gained from the expert advice of the Literary and Linguistic Computing Centre in the University of Cambridge, the Oxford University Computing Service, and the Computing Centre of the University of Newcastle. The initial preparation of the data is the work of Julie Wade and Nicole Cox. In its analysis, I have been much assisted by the special skills of Alexis Antonia and the ingenious programs designed for me by Sandra Britz and David Hoole.

[2] Perhaps the most celebrated attack is that of Stanley Fish, in "What is Stylistics and Why are They Saying Such Terrible Things About It?", *Is There a Text in This Class?* (Cambridge, MA: Harvard University Press, 1980), pp. 68–96, 246–67.

[3] The methods are described in more detail in my "Word-Patterns and Story-Shapes: The Statistical Analysis of Narrative Style", *Literary and Linguistic Computing*, (1987), 61–70.

[4] *Ibid.*

[5] Duncan Isles, in Charlotte Lennox, *The Female Quixote*, ed. Margaret Dalziel, (London: Oxford University Press, 1970), p. xxii.