Vectors of Prose Style

JOHN B. CARROLL

The very concept of style implies variation. It takes little argument or evidence to secure agreement that there are different manners of writing, and that these differ among themselves not only by virtue of the content or the subject matter treated but also by virtue of a host of "stylistic" elements which are present in varying degree in samples of prose. But what, exactly, are these stylistic elements? Ever since man discovered the pleasure of commenting upon his own and others' oral and written compositions, he has been seeking a useful set of pigeonholes for classifying style. The tendency has been for the classifications to proliferate without design or system. Literary criticism today does not have any well and sharply defined set of elements by which a sample of prose may readily be characterized.

In 1935, the renowned psychologist L. L. Thurstone published a book **(414)** in which he presented the technique of what is generally known as *factor analysis*—a statistical procedure for identifying and measuring the fundamental dimensions ("vectors") that account for the variation to be observed in any set of phenomena. Since then factor analysis has been a tool widely used by psychologists in studying intelligence, personality, interests, emotions, rates of learning, and even word meanings, but the technique has never heretofore been applied to the study of literary style. If we can study the "personalities" of people by factor analysis, we should be able to study the "personalities" of samples of prose. In the simplest possible terms, factor analysis enables the investigator to apply a large number of measurement procedures to a sample of objects and find out to what extent these measures overlap with each other.

Although the objective study of literary style by means of statistical analysis is not a completely novel endeavor, none of the scholars who have engaged in such study has ventured to ask the question raised here: what are the basic dimensions in which style varies? In contrast to previous statistical studies of style, each of which has fixed attention on one or a

small number of the possible ways of measuring style, this investigation[1] examines the relations among a large number of indices of style and attempts to identify the most salient ways of describing stylistic variation in prose.

The notion of attempting to quantify aspects of literary style will be repulsive to many literary critics and outright ridiculous to others. The writer must confess that even he, after completing the study, remains skeptical whether the dimensions identified here adequately represent the aspects of style that truly make the difference between great literature and the not so great, or even the aspects that serve to differentiate some of the recognized styles of writing. Nevertheless, some of the hopes in which the study was undertaken seem to have been realized: the study points to some of the more obvious characteristics of prose which have to be observed, mentioned, and duly noted before the literary critic can really go to work. It injects a semblance of order into the study of "readability" and suggests certain bases for guiding the teaching of English composition in schools. Further, it provides leads toward the psycholinguistic study of the "encoding" processes by which the individual translates nonverbal prelinguistic states of behavior into linguistically encoded output. It lends some support to the notion that certain factors of literary style correspond to predispositional "sets" which govern the emission of large classes of verbal responses—personal pronouns, for example.

PROCEDURES

There are two distinct kinds of problems to be faced in designing any study that seeks to identify the major dimensions of a set of phenomena: (1) how can we obtain a sufficiently heterogeneous sample of the things we want to study, and (2) what measurements shall we take in order to sample

all the significant ways in which the phenomena vary? Practical considerations set certain limits in both of these problems.

The sample of objects studied here consisted of 150 passages from various sources and styles of English prose. Each passage was chosen so as to be more or less self-contained within a little more than 300 words. By selecting passages according to categories—novels (both British and American, both nineteenth and twentieth centuries), essays, newspaper features and editorials, biographies, scientific papers, textbooks, speeches, legal documents, personal letters, and sermons were among the categories used—we hoped to include the widest possible assortment of subject matters and styles. The sample even included several relatively low-grade high-school English compositions.

The measures taken on these 150 passages fell into two classes: subjective and objective. The objective measures involved various counts, indices, and ratios based on the enumeration of certain classes of words, clauses, sentences, and other linguistic entities and included some of the measures used in previous statistical studies of style. Subjective measures were secured partly to help in the interpretation of results for the objective measures, partly to provide bench marks for certain characteristics of style which the objective measures could hardly be expected to describe. It was of intrinsic interest, also, to study the extent to which a group of competent judges could agree in assigning ratings, and to determine the totality of ways they could find for characterizing the passages. In order to make the rating task as simple as possible, 29 adjectival scales were chosen with a view to covering the major qualities and traits of style as far as they could be determined a priori, and 8 expert judges—all with interest and training in English literature—were secured to rate each of the 150 passages on each of the 29 scales, the form of which may be illustrated as follows:

meaningless ____ : ____ : ____ : ____ : ____ : ____ : ____ meaningful

The 8 judgments obtained for each passage on each scale were then averaged.

In all, 68 scores were obtained for each of the 150 passages: the 29 averaged ratings of the 8 judges, and 39 objective measures. The names of the measures are listed in the first column of Table 1; unfortunately, space does not permit a full description of the procedures for obtaining the objective measures. The resulting 68 × 150 scores formed the basis for the ensuing statistical analysis. The correlation of each measure with each other measure was determined—the results being exhibited in a very large table with 68 rows and 68 columns. This *correlation matrix* was then subjected to a factor analysis in order to determine how many fundamental

Table 1. **Results for 68 Measures of Prose Style; Reliability Coefficients and Loadings on Six Dimensions (Factors) of Prose Style**

Variable	Variable Number	Reliability	General Stylistic Evaluation A	Personal Affect B	Ornamentation C	Abstractness D	Seriousness E	Characterization F
*Subjective Ratings**								
Profound-superficial	1	.84	.43	-.11	.06	.53	.41	.15
Subtle-obvious	2	.81	.20	-.17	.09	.72	-.09	.15
Abstract-concrete	3	.90	.02	-.01	.18	.64	.11	.16
Meaningful-meaningless	4	.70	.70	-.11	-.04	.04	.41	.03
Succinct-wordy	5	.78	.51	-.15	-.65	-.15	.14	-.01
Graceful-awkward	6	.73	.84	-.01	.07	.17	-.08	.12
Vigorous-placid	7	.80	.26	.63	.17	-.06	.21	-.10
Lush-austere	8	.80	-.01	.43	.55	.07	-.29	.17
Earnest-flippant	9	.87	.06	-.05	-.01	.05	.71	.13
Intimate-remote	10	.87	.10	.82	.02	-.33	.03	-.01
Elegant-uncouth	11	.82	.33	-.29	.44	.40	-.11	.02
Natural-affected	12	.80	.49	.06	-.51	-.26	.29	.08
Clear-hazy	13	.78	.72	.16	-.09	-.45	.10	-.17
Interesting-boring	14	.78	.84	.25	-.04	.12	.01	.06
Strong-weak	15	.64	.88	.21	.13	.00	.11	-.07
Opinionated-impartial	16	.89	-.01	.53	.36	.01	.07	.02
Original-trite	17	.77	.54	-.08	.04	.44	-.20	.15
Ordered-chaotic	18	.69	.65	-.29	-.04	-.15	.06	-.09
Vivid-pale	19	.80	.61	.54	.07	-.09	-.05	-.01
Personal-impersonal	20	.86	.03	.83	.14	-.26	.00	.01
Precise-vague	21	.71	.64	.00	-.10	-.47	.03	-.17
Masculine-feminine	22	.85	.22	.06	.08	-.08	.58	-.09
Varied-monotonous	23	.75	.75	.22	.04	.20	-.10	.15
Emotional-rational	24	.90	-.05	.77	.22	-.12	-.02	.09
Complex-simple	25	.82	-.09	-.18	.48	.51	-.07	.04
Pleasant-unpleasant	26	.75	.88	.15	-.03	.07	-.11	.08
Serious-humorous	27	.92	.01	-.12	.00	.13	.70	.08
Florid-plain	28	.82	-.08	.33	.66	.16	-.28	.11
Good-bad	29	.74	.95	.01	-.02	.12	-.05	.02

Objective Measures (All based on 300 words)								
No. paragraphs	30	—	—.09	—.13	—.50	—.02	—.23	—.03
No. syllables	31	.86	—.10	—.58	.30	.05	.09	—.23
No. sentences	32	.85	.05	—.03	—.61	.13	—.18	.05
Standard deviation sentence length	33	.17	.11	.25	.54	.00	.12	—.01
No. clauses	34	.78	.01	.28	—.60	.26	—.22	.04
Clause complexity index	35	.60	—.08	.18	.39	.02	.18	—.15
% noun clauses	36	—	—.02	.01	—.22	.45	.13	—.25
% adjectival clauses	37	—	—.13	—.07	.19	—.21	.05	.34
% adverbial clauses	38	—	.11	.09	.04	—.20	—.17	.06
% parenthetical clauses	39	—	—.01	.04	.08	—.01	—.02	—.29
% "action" verbs	40	.76	.06	.11	—.40	—.17	—.23	—.01
% "cognitive" verbs	41	.58	.06	.46	.07	.12	.05	—.06
% transitive verbs	42	.33	.01	—.06	.03	—.15	—.03	—.63
% intransitive verbs	43	.26	—.14	.14	—.09	—.05	—.17	.32
% copulative verbs	44	.44	.04	—.04	.06	.18	.12	.43
% Latin-derived verbs	45	.57	—.06	—.33	.20	.12	.10	—.25
% passive verbs	46	.49	—.06	—.45	.15	—.07	.16	.12
Mean tense	47	.88	—.12	—.15	.11	—.02	—.01	.04
Entropy of tense	48	.42	—.16	.07	—.02	.06	.04	—.29
No. infinitives	49	—	.04	.16	—.14	—.01	.05	.02
No. participles	50	—	.08	—.08	.17	.31	—.07	—.15
No. gerunds	51	—	.13	—.20	—.03	.08	—.04	—.15
No. proper nouns	52	.83	—.04	.01	—.03	.01	—.01	—.38
No. common nouns	53	.74	—.01	—.49	.18	—.22	—.11	.13
% unmodified common nouns preceded by "the"	54	.45	.07	.07	—.50	.10	.03	.18
% nouns with Latin suffixes	55	.69	—.06	—.22	.42	.22	.20	—.22
No. articles	56	.52	.18	—.28	—.11	.01	.04	.20
% indefinite articles	57	.48	.07	—.09	—.10	—.01	—.48	.01
No. personal pronouns	58	.84	.11	.60	—.23	.20	—.10	—.09
No. possessive pronouns	59	.64	—.07	.30	.09	.08	—.21	—.14
No. indefinite pronouns	60	.43	.01	.28	—.18	.14	.21	.27
No. indefinite and quantitative determiners	61	.27	—.10	—.09	.11	—.24	.35	.20
No. demonstrative pronouns	62	.44	—.18	.11	—.05	.05	.22	.13
No. numerical expressions	63	.67	—.01	—.07	—.09	—.49	—.06	—.02
No. prepositions	64	.58	.01	—.34	.35	—.01	.09	—.01
No. pronouns	65	.83	—.01	.58	—.26	.18	—.14	—.13
No. determiners	66	.52	—.19	.01	.00	—.45	.34	.25
No. descriptive adjectives	67	.66	.11	—.18	.38	.09	—.10	.20
No. participial modifiers	68	.38	.09	—.05	.33	.01	—.13	—.01

* The positive pole of the adjectival scales is represented by the *first* of the two adjectives specified.

dimensions would be needed, at a minimum, to account for all the inter-relations among the 68 measures.[2]

RESULTS

The formidable appearance and size of Table 1 are due simply to the desire to compress a maximum of the essential results of the study into a single table; the reader is invited to examine it carefully. It contains information bearing on two kinds of questions about the 68 measures. (1) How "reliable" are the measurements? In the case of the 29 subjective measures, this question relates to the extent to which the judges agreed in their ratings. A reliability coefficient of 1.00 would denote perfect agreement, and a coefficient of .00 would denote purely random agreement. In the case of the 39 objective measures, the coefficients given in Table 1 (where they are present at all) refer to the extent to which each measure gives consistent results from the first half of a 300-word sample to the second half. (2) What general trait or traits does each variable measure and to what extent? The data relevant to this question are the coefficients found in the last six columns of the table. All coefficients larger than about .25 in absolute magnitude may be regarded as significant for purposes of interpretation.

The reliability coefficients (in the first data column of Table 1) for the 29 averaged subjective ratings range from .64 for the scale *weak-strong*, to .92 for the scale *humorous-serious*, with a median at .80. Although the figures are high enough to suggest that each measure is sufficiently reliable to give meaningful results, the lack of perfect agreement is particularly noticeable for some scales. Some scales, such as *meaningful-meaningless* and *ordered-chaotic*, have low reliability because, we may guess, judges differ in their conceptions of how these terms apply to prose passages. It is of more than passing interest that scales which (as will be seen later) denote general stylistic evaluation, such as *good-bad, pleasant-unpleasant, strong-weak, interesting-boring, graceful-awkward, varied-monotonous, clear-hazy*, have uniformly low reliabilities, whereas such scales as *serious-humorous, abstract-concrete, emotional-rational, opinionated-impartial, earnest-flippant, intimate-remote*, and *personal-impersonal*, all of which refer to specific and relatively nonevaluative qualities of style, have high reliabilities. Judges can often agree in making descriptive classifications

---

[2] Of possible technical interest to some readers is the fact that the initial factor analysis was performed by means of Thurstone's centroid method, after which the factors were "rotated" to oblique simple structure by the writer's so-called *normal biquartimin* criterion (57). All these computations were performed with the aid of high-speed electronic computing machines.

of prose passages but they agree less often in making general evaluations of style. Perhaps this is what makes literary criticism exciting.

Concerning the reliabilities of the objective measures, we shall only comment that the figures indicate the extent to which a writer is likely to hold certain formal characteristics of his style constant within relatively short stretches.

We come now to the main findings of the study, the findings that give a provisional answer to the question of what are the dimensions of literary style. Although *seven* dimensions were indicated by the factor analysis technique, it appeared that only six of these could be given meaningful interpretation, and thus the data for the seventh are omitted from Table 1. The order in which the six remaining factors are discussed is actually immaterial, but they are listed in Table 1 as factors *A*, *B*, *C*, *D*, *E*, and *F*, in order of their apparent interest, importance, and relevance in connection with the study of literary style.

The variables having high coefficients in column *A* of Table 1 are in every case subjective ratings. In order of the magnitude of their "loadings" (as the coefficients are often called) they are the scales *good-bad* (29), *pleasant-unpleasant* (26), *strong-weak* (15), *interesting-boring* (14), *graceful-awkward* (6), *varied-monotonous* (23), *clear-hazy* (13), *meaningful-meaningless* (4), *ordered-chaotic* (18), *precise-vague* (21), *vivid-pale* (19), *original-trite* (17), *succinct-wordy* (5), *natural-affected* (12), *profound-superficial* (1), *elegant-uncouth* (11), and *vigorous-placid* (7). All these scales, in differing degrees, denote over-all positive or negative evaluation of a prose passage. We are therefore inclined to identify this factor by the name General Stylistic Evaluation. Notice, however, that some of the scales have significant loadings on certain other factors. Only the first six scales mentioned are unequivocal measures of stylistic evaluation alone. It is cheering to note that not a single objective measure shows any significant loading on factor *A*, General Stylistic Evaluation. Although the style of literary passages can be indexed in certain ways mechanically, it cannot be *evaluated* mechanically!

The key to the interpretation of factor *B* seems to be the presence of the subjective scales *personal-impersonal* (20), *intimate-remote* (10), *emotional-rational* (24), *vigorous-placid* (7), and to a lesser extent *vivid-pale* (19) and *opinionated-impartial* (16). Let us call this dimension Personal Affect. It is also indexed by a number of objective measures, such as number of personal pronouns (58), number of pronouns (65), and (negatively) number of syllables (31). (The negative loading of number of syllables is to be interpreted as indicating that passages with high Personal Affect have a relatively small number of syllables in 300 words, that is, the words tend to be short.) The dimension of Personal Affect is unrelated to General

Stylistic Evaluation: it refers simply to the extent to which a passage uses personal references, emotive terms, and similar devices, without necessarily making for "good" style or for "bad" style, either, for that matter.

Let us proceed to column *C* in Table 1. If the reader will run his finger down this column he will find high loadings for the following subjective scales: *florid-plain* (28), *wordy-succinct* (5) [reversing the polarity of the scale makes the loading positive], *lush-austere* (8), *affected-natural* (12), *complex-simple* (25), and *elegant-uncouth* (11). The factor is also indexed by long sentences (measure 32), long clauses (34), wide variation in sentence length (33), a relatively high proportion of common nouns which are preceded by adjectival or participial modifiers (54), long paragraphs (30), a high proportion of nouns with Latin suffixes (55), a low proportion of verbs denoting physical action (40), a high degree of use of dependent clauses of various orders (35), and a high number of descriptive adjectives (67). "Ornamentation" (as opposed to "plainness") is clearly a suitable name for this dimension.

The subjective scales having high loadings on factor *D* are *subtle-obvious* (2), *abstract-concrete* (3), *profound-superficial* (1), *complex-simple* (25), *hazy-clear* (13), *original-trite* (17), *elegant-uncouth* (11), and *remote-intimate* (10). The common element in these scales seems to be a generalized notion of abstractness and obscurity as opposed to concreteness, precision, and perspicuity; for convenience let us call this dimension Abstractness. Like factors *B* (Personal Affect) and *C* (Ornamentation), it is independent of factor *A* (General Stylistic Evaluation); that is, abstractness versus concreteness, the use of personal references versus the failure to use them, and ornamentation versus plainness have nothing to do with whether a prose passage is favorably thought of or with each other. Factor *D* (Abstractness) can be fairly well measured by several objective indicators: by a low proportion of numerical expressions (63), a low number of determining adjectives and pronouns like "this," "each," etc. (66), a high proportion of noun clauses (36), and a low number of participles (50).

Factor *E* we call Seriousness. The two subjective scales measuring this factor best are *earnest-flippant* (9) and *serious-humorous* (27). We are somewhat surprised to find, however, that the scale *masculine-feminine* (22) also relates to this factor. Evidently the term "masculine" as applied to literary style connotes earnestness and seriousness, whereas flippancy and humor are associated with femininity. Other scales measuring seriousness are *meaningful-meaningless* (4) and *profound-superficial* (1), and the factor can be indexed objectively by a low proportion of indefinite articles (57), a high proportion of indefinite and quantifying determining adjectives (61), and a high number of determiners (66). Whether these objective measures are intrinsically related to seriousness, or whether the findings are simply a

reflection of the particular sample of literary passages used, we do not know.

Factor *F* in Table 1 is measured exclusively by objective measures: a low proportion of transitive verbs (42), a high proportion of copulative verbs relative to all verbs (44), a low number of proper nouns (52), a high
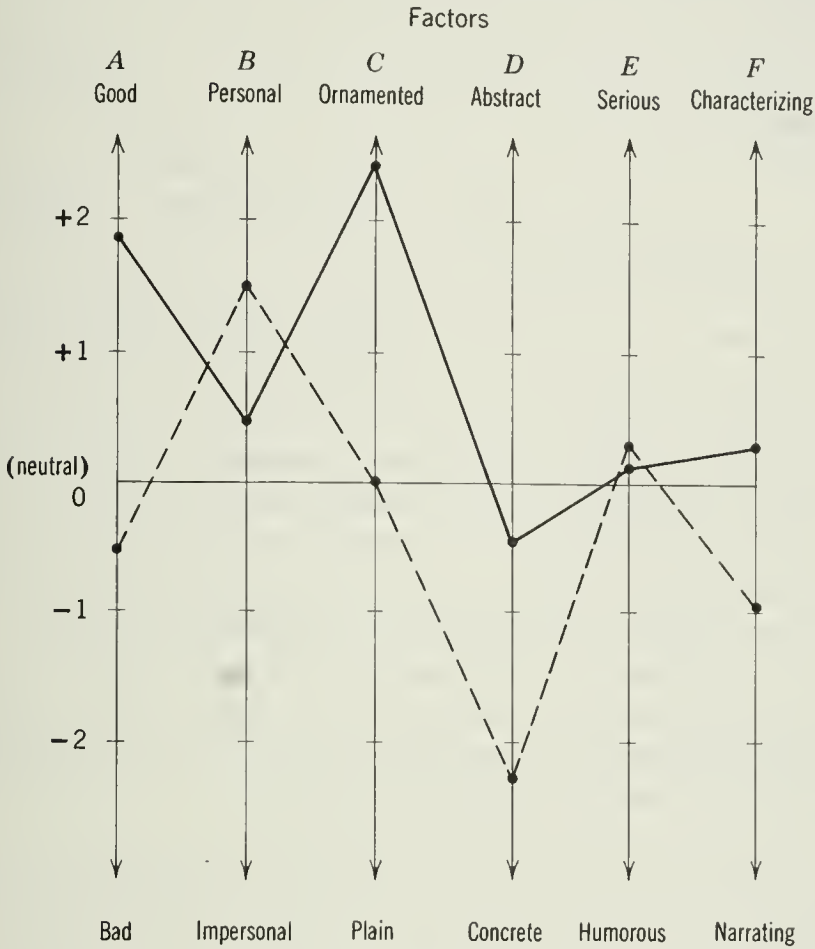

Figure 1. Style profiles of two prose passages: (———) a selection from F. Scott Fitzgerald's *A Diamond as Big as the Ritz*; (– – – –) a selection from Mickey Spillane's *Vengeance is Mine*.

proportion of adjective clauses (37), and a high proportion of intransitive verbs (43). We can make only a tentative interpretation of the underlying significance of this dimension; the evidence seems to point to a dimension of Characterization versus Narration. We would expect passages with high scores on this factor to be those that are more concerned with the "characterization" of entities—either by equating them with other entities through the use of copulative verbs or by describing them through the use of adjective clauses. Passages with low scores on this factor are more likely to be concerned with the reporting of action, most frequently the

action of persons; they would thus be found to have a high proportion of transitive verbs and proper nouns.

These, then, are the six independent dimensions of "style" which have been identified in this study: General Stylistic Evaluation, Personal Affect, Ornamentation, Abstractness, Seriousness, and Characterization versus Narration. Just to intimate the possibility of using these dimensions as the basis for a typology of style, we present in Figure 1 the "profiles" of two of the passages measured in the study. One was a selection from F. Scott Fitzgerald's short story *A Diamond as Big as the Ritz*, in which the author paints a vivid picture of the impressions of the hero and his companion as they wander through the diamond palace. The other is a passage from a very different sort of writing, that of Mickey Spillane.

Two questions may have harassed the reader: are all these dimensions really of "style"—are not some of them rather a matter of the content of a passage? And are not some of these "dimensions" merely dimensions of the *meaning* of adjectives, not necessarily dimensions truly inherent in samples of prose? With respect to the first question, we must reply that there is no hard and fast distinction between style and content. Try as we may to define style as the *manner* of treating subject matter, the type of subject matter will in general impose constraints upon the possible kinds of stylistic treatment. In the present study a vain attempt was made to have judges differentiate between content and style: *content* was to be rated with scales 1 through 4, and *style* was to be rated with scales 5 through 29. The results make it abundantly clear that the judges did not differentiate content and style, at least not in their ratings.

With respect to the second question, it must be insisted that even though somewhat comparable results might be obtained by asking raters simply to judge adjectives for similarity of meaning in the abstract, the dimensions of meaning themselves cannot exist without some support in the real world to which the adjectives presumably refer. The reality and substantiality of these dimensions is further attested to by the abundant instances of correlation between subjective ratings and completely objective, quantitative indices which derive meaning only when applied to actual samples of prose.