CURRENT TRENDS IN STYLISTICS

Edited by

BRAJ B. KACHRU
and
HERBERT F. W. STAHLKE

published by LINGUISTIC RESEARCH, Inc.

MODELS, COMPUTING, and STYLISTICS

Sally Yeates Sedelow
University of Kansas

&

Walter A. Sedelow, Jr.
University of Kansas

By way of general orientation it may be useful to suggest that style-content dichotomies seem to have been heavily impedimental to intellectual movement in the direction of an adequate stylistics. Per contra, it would seem very much in order to subsume within a comprehensive stylistics all discussion of content. Both from the immediate experience of work directed toward the exact characterizing of very long language strings and from the philosophy of scholarship and science, it would appear clear that theory integration is highly desirable in a concept set that is to be used post-heuristically, used in work meant to be demonstrably cumulative in character and impact. Studies of the literary sub-set of language usage traditionally have been marked by a lack of explicit precision in method--by, that is, a lack of operational definitions and, necessarily then, a lack of appropriately tight conjunctions between theory elements and such definitions. Stylistics probably cannot break out of that vague circularity until operational definitions, replicable investigations, concept integration, and predictive validation are well-established characteristics of the investigative style of students of literature.

While, in principle, those characteristics could be aspects of the method of investigators using style-content dichotomies, that could happen only if separately precise and internally coherent metalanguages of style-study and content-study were both mutually exclusive and, jointly, analytically comprehensive when fitted together under some new and more abstract rubric. Despite such a possible rough analogue as the history of wave/particle theory and research in physical science, as a research strategy in stylistics such a divided effort doesn't seem worth the price--even though it might have the consolatory value of 'saving' some traditional terms, albeit, ironically, doing so in a neologistic mode. What appears to be much the more usual case in science is that an older set of terms falls into neglect when new, more effective types of precision come into use. The relative lack of rigor and comprehensiveness vis-a-vis their phenomenal domains of style study and practice and content study and practice when taken either severally or jointly may lead us to expect those sets of terms (or, more exactly, one set and the current usage of the other) to deteriorate from their present usage of the other) to deteriorate from their present usage into historical curiosa. In any event, in time we will be able to tell how well their present users can sustain them and, also, how effectively other terminologies and procedures are employed.

Further, by way of additional orientation, it should be helpful to introduce here at least an intimation of the probable future lines along which theoretical rapport is likely to be achieved between formal information/communication theory and stylistics. The conceptual relationships of entropy and redundancy with style are likely to be of great moment, for communication research as well as for stylistics. Trying for the nonce to parallel traditional distinctions between content and style, one might say (with the important exception of the formation of new words) content is any choice of words from among the total ensemble of available words--so that, information-theoretically, we would derive from content the surprise or informational values for a literary passage and, cybernetically, its directive values within a fixed range. And style, then, is that management of redundancy implied by the particular string(s) assembled. Looking at the inherited semantic dowry stylistics brings to a union with the communications engineering sorts of research, it seems self-evident that content-free definitions of information are not going to be satisfactory for the broad band of communications re-

searchers; that may seem to be even more the case as, putatively, symbolic behavior becomes an increasingly central subject of research attention, however adequate such content-independent definitions and simplifications may have been, and may be, for a certain class of engineering studies. As the definitional structure of information approaches more nearly the comprehensiveness of the variable-sets informally associated with the parameterizing of real-world communication processes, we may expect not only that the selection processes will be approached with more than entropy measures, but also that those unique patterns of redundancy characteristic of particular chosen-word strings--patterns that, in a traditional sense of redundancy (narrower than that we employ), constitute what is intimated by stylistics traditionally--will come to be understood within the same taxonomic and predictive model as entropic definings of the content selection process. Here, then, might be the juncture at which style-content dichotomizings would fade from all but historical interest. At present abstract control-system theory, cybernetics, and general systems theory are all disadvantaged by a lack of coherence among the concepts of order, information, entropy and thermodynamics, even though when taken separately in context these term sets are clear enough. Some of the desired integration of formulations as between entropy and redundancy when applied to the problems of style (in the extended sense in which it is used in this discussion) prospectively could emerge as a consequence of this reconciliation of concepts as to order-disorder within the physical and symbolical sciences.

* * *

Turning a bit from this somewhat abstract discussion of theoretical issues and their implications, turning from some of the properties of a good model, to aspects of current work employing the computer as instrumentation in stylistics research, it must be acknowledged that there are many different points of vantage from which to discuss computing and stylistics. The definition of style used in our own research is this: that style consists of the patterns formed in the linguistic encoding of information. We first produced this definition in 1963. Recently we noticed that in 1958 John Carroll was using a highly similar formulation for his discussion of vectors--or characteristics--of prose style. Carroll argued that identification of such characteristics would provide 'leads toward the psycholinguistic study of the "encoding" processes by which the individual translates non-verbal pre-linguistic states of behavior into linguistically encoded output.' (1969:148) Our definition was meant to facilitate interaction between style research and the results of interest in trying to discover mappings between those languages humans use for communication with other humans and those used by the human central nervous system.

One vantage point for discussing computing and stylistics is simply to indicate the ways literary scholars are using the computer, in the hope that others not only will make extrapolations for their own fresh uses of the computer, but also will subsequently derive from such uses conclusions apropos the theoretical framework structuring their research and teaching. But in this main section of this essay we will devote attention explicitly to that theoretical framework, and specific uses of the computer will be mentioned only to illustrate and clarify more general remarks. Here we shall first explore somewhat further the relationship between the uses of the computer and a science of stylistics and, second, discuss the relationship between a science of stylistics and literary research and teaching.

Sometimes literary scholars and students assume that the use of the computer signifies the intrusion of science into literature, or, more broadly, into the humanities. Historically, literature has derived much inspiration from science and technology and, as will become plain, we do not regard science as an 'intruder.' Of course it is false to equate the use of the computer with an introduction of science--except for the fact, in this case trivial, that the computer itself has resulted from science and its technological applications. Just as playing with one's toes--which could be regarded as rather primitive, digital computers--need

not imply scientific computation, so, much activity involving an electronic digital or analog computer should not be termed scientific. Origins are one thing, applications another.

Crudely put, we may say that the components of a set of activities which might together be termed scientific would include (a) observation of a set of events; (b) the formulation of an hypothesis about the structure or behavior of what has been observed; (c) testing of the hypothesis through further observation so that the hypothesis may, if indicated, be judged inadequate, with the result that it is either rejected, or modified and tested again; (d) especially when organized into a model, predicting with non-invalidated hypotheses the behavior or structure of other entities or events comparable with the observed entities and events when under somewhat altered contingencies. In his 'Polemical Introduction' to the Anatomy of Criticism, the well-known literary critic Northrop Frye uses such terms as coherence, unified structure of knowledge, systematic, and progressive when discussing the scientific approach. (1957:3-29) In recent years--coincident with the increase of academic attention to the philosophy of science (an activity which, incidentally, Marvin Minsky has called 'literary criticism')--the word model often has been used in conjunction with hypotheses, as a component in the processes of scientific inquiry. Conventional dictionary definitions of model reveal at least two major uses of the notion of modeling, each of which can be applied to the study of stylistics. One definition describes a model as a miniature representation of a thing; sometimes, a facsimile. The second definition is that a model is 'a pattern of something to be made'. The first definition might be associated with a descriptive or even taxonomic approach to the study of style. That is, one examines prose or poetry, isolates elements in the text being examined and uses them to try to, in a sense, build a (partial) description, or verbal re-presentation, of some of the aspects at least, of the text being discussed. But short of the action of prediction, ultimately we have no epistemologically sound tests of descriptions, and it is a merit of the second definition of model that it might be associated with a generative approach to the study of style. A generative approach implies that one has created the pattern, template, analysis, or model, and in order to test it will try to generate (sample) equivalents of the thing being modeled. In scientific practice, the descriptive approach and generative approach may represent alternate sides of a coin. If the material produced by synthesis, that is by the generation from the model, does not meet one's criteria for satisfactory closeness in characteristics to that being studied, then the model must be changed. We would say, then, that the construction of a descriptive model of literary language strings is related to the observational and analytical components of science, and the construction of a generative model in some sense parallels the use of hypotheses, etc. in the process of synthesizing whatever is being modeled. Clearly, if one ultimately succeeds in constructing a good model, what one has is a model that is so descriptively precise that it can be used to generate or produce the observed event or entity being described. In the setting of literary scholarship one could have a model used to produce the style being studied.

A major advantage in using the notion of model--especially of a generative model employed to produce aspects of a specified style--is that it forces one to pay attention to the inter-relationships among elements--for example, to their complementarities, to their synonymies and antonymies. In the introductory chapter to her book, Style and Proportion, The Language of Prose and Poetry, Josephine Miles (1967:5) notes that ' "a characteristic passage" in prose must mean not just any section of given length, inasmuch as we should not expect or wish for homogeneity of character throughout a text, but rather a "passage", a progression which does actually move forward by a major predicate in ways, with segmentations or subordinations, representative of those habitual in the text when it is making such progress.' By segmentation, Miss Miles means a text which is explicitly, clearly sequenced by the grammatical structure, e.g., 'God who is invisible has created the world which is visable.' By subordination, she implies a more succinct phrasing, e.g., 'Invisible God has created the visible world' (1967:4).

> A whole sentence or a whole long paragraph may take one main step: I came; and this step may be expanded as to where, why, or how. It may take three: I came, I saw, I conquered; where, when, why, how... For these reasons of assumption [i.e. the cutting away of predicates and replacing them with adjectives as in "Invisible God" for "God who is invisible"] and subordination, these powers of prose to expand and contract and shift its emphases, the criterion of length of word, sentence, or paragraph is an especially irrelevant one, much as it has been used by critics of prose style. Is a short word more economical than a long one or must three be used as equivalent? Is a sentence that spares verbs more economical than one that spares assumptions? ...At least we may say, therefore, that a characteristic passage needs to be recognized in terms of a text's whole style of progression or its general habits of little leaps, heavy landings, or long articulated strides. (1967:5)

The point we wish to extract from Miss Miles' remarks is the emphasis upon the 'text's whole style...Its general habits.' Through good modeling we may perhaps begin to move toward the complexity requisite to the delineation of the general habits of a complete text.

At this point one might ask whether a computer is needed for models. The answer, of course, is no. There were models of various sorts long before the advent of the computer. An example might be Sidney Lanier's (1891) attempt to study sound in poetry without reference to ideas by using a musical model which enabled him to talk in terms of musical notation about relationships among sounds. His model was not scientific; it lacked an important component of the scientific approach--replicability. That is, although poetry may be music to us all, one may doubt that all of us, or even two of us, hear the same melodies; in a manner of speaking we do not, one might surmise, hear the same absolute pitches and duration of sound nor, presumably, even the same relative shifts in pitch and duration. But, in any case, Lanier did not provide the operational definitions of terms requisite for establishing common referencings for his meanings; and the presence of that property is, in some sense, a sine qua non for science.

A model which is rather intimately allied with computing and which has seen some use for the study of language and literature is the classical paradigm in information theory. It seems appropriate to devote some attention to it at this point. We draw on Singh (1966) and Pierce (1965) for some particularities in the expression of these ideas. As is well known, and as indicated in the opening section of this essay, information theory, narrowly speaking, is associated with communications engineering. That is, it is concerned with efforts to transmit information with the least possible distortion, using such devices as the teletype, telephone, radio, and television set. The communications engineer stresses the quantitative aspect of the flow in a communication system of an intangible attribute called information. Information, in this context, is measured by its 'news' value, that is, the extent of surprise it causes the recipient. The importance of this definition for a communications engineer is clear if you consider the engineer's concern for delivering a message to the receiver in a form as close as possible to that in which it left the sender, or transmitter. Thus, he needs a measure which will enable him to determine how well he's succeeding at delivering the message. The evaluation of his performance can quite easily be muddied by cleverness on the part of the receiver which, or who, can use redundancies or patterns of various sorts to help decode the message, so that the sender might--wrongly--conclude that the message was well received because it had been transmitted so clearly through the communication channel. Therefore, in order to get a clearer picture of the effectiveness of the system, the engineer needs a measure of what's new, what's unexpected, in the message received. It should be emphasized that the 'newness' or 'unexpectedness' has nothing to do with semantic content in everyday usage--rather, it is concerned with, (1) whether or not a

possible message element occurs or doesn't occur (notice that two states, or conditions, are involved here) and, (2) whether the element occurs in a message sequence with sufficient regularity to be predictable (that is, what is the probability of that element occurring?). Another way of thinking about probability of occurrence is suggested by J. Singh (1965:12-3) when he says that 'information in this context is merely a measure of one's freedom of choice when one selects the message from the available set, many of which may well be devoid of meaning.'

These considerations are reflected in the formula for the average information content of the message:

$$- (P_1 \log_2 P_1 + P_2 \log_2 P_2 + \ldots + P_n \log_2 P_n) \text{ bits}$$

This formula should be read as: The probability of message event (or message) one times the log to the base 2 of the probability of message event one, plus the probability of message event (or message) two, etc. Bit is a contraction from binary unit or binary digit, referring to <u>two</u> possible message states--occurrence, nonoccurrence; on-off; dot-dash, 1/0, and so on. Thus, the average information content of a one event message with a probability of 1/2 is:

$$- (1/2 \cdot (-1)) = 1/2;$$

minus 1 comes from ($\log_2 1/2$), which is to say: 2, raised to what power, equals 1/2. That is: $2^{-1} = 1/2$. (A negative exponent always equals 1 over the base, raised to the positive power of the exponent.)

Notice that, as you take the log to the base 2 of probabilities, the <u>absolute</u> value of the log increases as the fraction grows smaller. For example the log to the base 2 of 1/8 = 3 (in absolute terms); the $\log_2$ of 1/16 is 4, and the $\log_2$ of 256 is 8. Now, in fact, the $\log_2$ of 8 is -3, and the log 2 of 16, -4, and so on. But what happens to that negative sign? It is multiplied by a negative sign to produce a positive value; neither Singh nor Pierce point out that the final positive value is part of the mathematical model for information content; rather, they both simply say in effect that 'the negative sign at the beginning of the expression is added to make the expression positive since, as it happens, logarithms of all proper fractions like 0.9 and 0.1 are negative numbers.' In returning to a discussion of the logs of probabilities, you can see that as the probability of an event declines, its information content (the information content of a message event is the $\log_2 P_n$ part of the formula) increases. Now, to get the <u>average</u> information content for a message comprising more than one event, the information content of each event is multiplied by its probability of occurrence. As you know, probability is always expressed as a fraction and when any number is multiplied by a fraction the value of that number is reduced. So that, multiplying by the probability weight <u>then</u> reduces the information content of a small probability so as to keep the resulting measure of average information content for the message in line with the fact that although an event with small probability carries a lot of information, <u>it doesn't occur</u> very often. Therefore, its expected information content is lower.

Now, given that perhaps unnecessarily extended layman's introduction to the formula/statistical model expressing information content, just how useful is that model as a measure of information and, in a more extended sense, as a method for examing style? For communication engineers, information theory has been quite useful, providing them with a measure on which to base efforts to achieve reliable transmission. For those interested in information in a broader sense, this approach to the definition of information has not been so useful. For example, if one considers all the books in a library a message comprising a very large number of message events, the prospect to a student or scholar of receiving that message in a form as close as possible to that in which it exists is not at all cheering. Further, the prospect of having the information content of this enormous message evaluated without regard to semantics is alarming. You, of course, can quickly respond that it is unfair to consider the entire contents of a library as one message. Rather, one ought to sub-section, or classify, the library so as to

consider as the message only those references relevant to the given user's needs. But how, using information theory, could one select that relevant sub-set? The answer is that you couldn't, again because, to date at least, information theory has ignored semantics. If one had the conventional tireless computer and all the works in the library in computer-accessible form, one could provide the user with measures of information content of individual words, or groups of words, or punctuation marks, or whatever can be specified as message events within the message whole, and the user could, if he chose, say that everything exceeding a certain threshold was relevant to his topic. But clearly the material produced would range over the coverage of the library and might well have nothing to do with the information requirements of the individual users. Perhaps this sort of exposition helps to explain why people concerned with information retrieval are anxious to find some other way of defining and measuring information. And perhaps this Shannon measure could be used in a theory development that would deal with meaning as well as occurrence. It is encouraging to note, for example, that in a perhaps vaguely analogous situation, some modern linguists who earlier restricted their models of language to syntax are now beginning to attempt to incorporate semantics into these models. And perhaps, to return to the information theory model, one could argue that efforts to match terms employed by a user with those occurring in a set of documents or volumes in the library might be mapped onto some sort of expression employing a feedback loop between the verbal patterns in the user's message and those in the data bank he is querying.

Information theory has been used to investigate an aspect of literary style by the Russian, Kondratov. In a paper entitled 'Information Theory in Poetics: The Entropy of Russian Speech Rhythm', Kondratov examines rhythmic irregularities in Russian speech and the variability in the degree of uncertainty (entropy) in various types of texts, ranging from scientific and expository prose to classical verse meter. The text was hand-coded so as to indicate the number of syllables in each word and the position of the stressed syllable within the word. Having thus encoded the text, the probability of occurrence(the statistical frequency) of each rhythmic word type could be computed and, in turn, the entropy for the rhythm of each of the eight texts could be determined using Shannon's formula. In order to establish the rhythmic entropy per syllable for any given text, the entropy per word was divided by the average word length. 'Not surprisingly, the entropy value diminished successively from scientific prose--which had the highest rhythmic entropy per word--to the iambs of Lomonosov, whose verse demonstrated the smallest degree of rhythmical uncertainty both for the word and syllable,' (Kondratov 1969:114). To provide a sense of the relative value, the entropy for scientific prose was 3.94 and for the poetry of Lomonosov, 2.75 (that is, the figure for entropy per word). Kondratov then went on to look at entropy for word pairs, with much the same results. There would appear to be no theoretical obstacle to calculating entropy for word strings of any length, but the practical difficulties of computing the frequency for such strings are considerable. That is, simply recording the possibilities so as to keep track of frequency can impose a strain on computers--either on computer storage or on computer time. At any rate, it is clear from this example and the foregoing remarks on information theory that if whatever one wants to measure can be clearly specified, information theory can then provide a measure of the degree to which the text is patterned with reference to the characteristic being investigated. But the important decisions--such as, in the case of Kondratov, how to define rhythm so that the encoding of the text could be replicated by someone else--are, it is important to note, not within the scope of information theory as presently defined and used.

A broad range of statistical and mathematical models has been employed to look at various aspects of literary texts. In his attempt to identify the basic dimensions, or parameters, or categories for prose style, John Carroll (1969) used factor analysis to group together responses to texts by human judges--under such headings as abstractness, seriousness, and personal affect. In a study

more objectively based, on the occurrence of words in texts, Josephine Miles and Hanan Selvan (1966) subjected lists of words used by 17th century poets to factor analysis in an effort to see which words tended to occur together. The purpose of this study was to see whether poets who had been traditionally classed together (e.g., the Metaphysical poets or the 'Sons of Ben') tended to use similar groups of words. In a 1970 dissertation for the English Department at the University of North Carolina, John Smith used factor analysis to study patterns of co-occurrence among words he defined as representing images in James Joyce's Portrait of the Artist as a Young Man. He was then able to study which images occurred together, which never occurred together, which sometimes occurred together, and so on. Using the results of the factor analysis, he was able to show that at the moments Joyce identified as those of epiphany, many of the major images co-occurred. Both the Miles-Selvan and the Smith studies used the computer to perform the factor analysis.

A glance at the papers in Dolezal-Bailey (1969) shows that scholars have used statistics in describing sentence lengths, word lengths, word choice (how many words an author uses and how often he uses them), and many other string properties.

Other models include, potentially, the use of rings and of graph theory as models for the exploration of the inter-relationship of literary themes--and we have been using such approaches in our own computer-based research. In a sort of modern equivalent of Lanier, non-verbal visual models are being used to represent literary texts. Computer-produced graphs and other diagramatic drawings describe, for example, the co-occurrences of images, the discontinuity of images, and other text features. John Smith used such graphs to make the results of factor analysis of Joyce's Portrait more vivid.

Other scholars are using categories which may someday become elements in a model or help to form a larger model than that currently supplied by the category itself. These categories include, for example, syntactic and phonetic categories. Louis Milic (1967) has used syntactic categories for his work on Swift and Josephine Miles for her work on the language of prose and poetry. Incidentally, for the syntactic aspect of the work, Milic used the computer and Miles did not. Analytical categories of course also include elements associated with myths, as when employed by Northrop Frye and others in exploring character types, narrative patterns, and overall themes; as noted earlier, the computer is being used for work with patterns of themes (e.g., folkloristic themes) and of imagery, and it is possible to imagine that out of such approaches, especially when taken for many different texts, will emerge statements concerning narrative patterns and character types. Other analytical categories include elements associated with literary history--for example, links between events in literary works and events in the 'real world'. (Is a given character patterned on a specific human being?) The use of the computer to detect such similarities would seem difficult, although insofar as the detection of those similarities depends upon the occurrence of words the computer might have utility. All of these analytical categories and approaches represent either a start toward a delineation of a model (the graphic representations function in this way) or the use of a model of some aspect of literary style (such as syntax) which might then be incorporated into a more comprehensive description--a description based upon observations, used in forming hypotheses, which are employed in constructing a model which can be tested. In other words, such efforts represent a move toward a science of stylistics.

The relevance of the computer to the building of models of style is that (i) it can 'perform' the observations which are tedious, drawn-out and therefore often ignored by the human being, (ii) it can cope, explicitly and systematically, with the many different stylistic variables, or elements, which are needed to generate a style according to the model provided, and (iii) in the process of generation, the computer's output can indicate to us the gaps in our observations and analysis, and suggest the ways in which the model should be modified. To take a couple of examples absurd to the point of clarity: if a model of English

included just adjectives and nouns, we would quickly see that the model didn't generate a string of idiomatic English, and we would refine the model; or, if the model dealt only with sentence lengths, the emptiness of the material generated would quickly be apparent.

All the models, whether they are mathematical or statistical or verbal (as in categories based on myth or historical literature) share one characteristic: they are bounded by the data and the definitions according to which the data is identified. Thus information theory offers a way to describe rhythmic patterning, providing one has a definition of rhythm and then applies it rigorously to data which has been replicably identified and collected. The mathematical or statistical approach has the advantage, relative to the verbal approach, of being coherent (that is, it can be more readily developed as a model), and of being replicable. Verbal models tend to be much less rigorously defined, and thus more easily tend to internal inconsistencies and shifts when employed by different scholars or even by one scholar at different times.

Statistical and mathematical models have been criticized because they deal with just parts of wholes--as, for example, information theory's obliviousness to semantics. Karl Kroeber (1969) makes similar points in his article 'Perils of Quantification: The Exemplary Case of Jane Austin's Emma'. Kroeber points out that simply counting word frequencies, for example, does not reveal the stylistic subleties of Emma. One would, of course, certainly agree, although one might not agree with his further remarks that 'it seems possible that any quantitative analysis of Literature destroys in the very process of analysis the literary qualities it seeks to describe' (1969:205). He goes on to argue that 'Quantitative analysis tends to make this fatal reduction; by discriminating parts it blurs the quantitative relationship of part to whole which in fact defines the essential nature of the part' (207). We agree that much quantitative analysis does make such 'fatal' reduction; in fact, perhaps, almost all of it to date exhibits this unfortunate piecemeal approach. Thus the major goal of research with which we are associated is to provide as comprehensive a picture as possible of style. To do so, we need to be able to identify multiple types of patterns and show their inter-relationships.

Our efforts thus far have been concentrated upon a few aspects of this undertaking: devising a system of computer programs for thematic (content) analysis; programs for graphic representation of thematic, or other patterns; and moving toward designing a 'package' of statistical programs which would count, measure and characterize anything which could possibly be so treated. The VIA programs in list-structured and ring-structured formats, as well as thesaurus research, are output from the efforts at content analysis. The MAPTEXT programs are for graphic pattern-representation. And a taxonomy of measures, as well as data formats permitting text-specific thesauri, verbal networks, syntactic description, etc., are for the statistical package.

We do not yet have a model of style, but we do have approaches which, we feel, are currently helpful for the interpretation of literature, for content analysis and other such applications; and we are beginning to move toward rigorous identification of parts which will we hope someday compose models of wholes.

Thus far, we have been avoiding the question implied in the promise at the beginning of this section to discuss the relationship between a science of stylistics and literary teaching and research. Perhaps some of you might ask, do we, indeed, want a science of stylistics? In partial support of an affirmative answer, consider this quote from Josephine Miles. She argues (1967) that 'analysis works to support and invite intuition. Often today we hear scepticism of the analytic because we ask it to do more than it can do. It does not create, invent, imagine, lead to values; but given values, it clarifies and discerns, helping us to understand the relation between what we feel and what we know.' We would be inclined to argue that analysis helps us understand the relation between what we feel and what we know; and one also might suspect and, certainly, hope that analysis does lead to values. One may be profoundly mistrustful of any

educational process which permits itself to be centrally based upon unexamined feeling or emotion. We think we understand some of the reasons for certain aspects of the current anti-intellectualism as expressed in the need to escape externally imposed rigor of various sorts; students may feel the need to escape, one might tentatively assume, because much of what is demanded of them seems senseless and, indeed, contradictory; it seems, therefore, irrelevant. One might also suggest that they may be moving away from what has been wrongly passed off as an analytical approach in the humanities because they discern that, in many ways, it is a fake. One can fully agree with Robert Gorham Davis' statement (1970:38), where interpretation and judgement are concerned, we can teach only opinions, not knowledge.' As we all know, and try as the New Critics might, we have been passing off as knowledge what was in fact impression and its derivation, opinion. But at the same time one may strongly disagree with the implications of the following remarks in the same article: 'If we are realistic about the limitations of criticism and the fact that when teaching literature we are dealing with opinions and sensibilities, not testing knowledge or truth, then we will be much freer to have our own opinions and to be true to our own sensibilities and to our own honest sense of justice or beauty' (38). We are not opposed to impressions and opinions and their concomitant feelings and emotions; and we are not opposed to candidly conveying those responses to students in a classroom or in the course of a scholarly article. But that is not to abandon knowledge and rationality as a humanistic goal. Earlier in his editorial, Mr. Davis states that such works as 'Susan Sontag's "Against Interpretation", Leslie Fiedler's "Waiting for the End" Marcus Klein's, "Beyond Alienation", and Lionel Trilling's "Beyond Culture" brought home to us the truth--still not fully acknowledged--that the teaching of literature is still not, and cannot be, a rational objective discipline like the teaching of science' (2). (Incidentally it is at least charitable to assume that Mr. Davis surely knows that there is much in science that is not rational or objective either.) In literature, as is in fact the case in science, we have room for both the emotional and the rational, feeling and objectivity. In literature we have not been objective in the past, because we did not know what to be objective about nor how to be thoroughly objective, and did not have the science and technology for doing so. Indeed, what student or scholar today is interested just in the length of sentences or in the number of words per sentence or paragraph? One is interested in such matters only insofar as they form part of a model which will enable the individual scholar and teacher and any other scholar/teacher/student to identify characteristics of a text which he can then relate to his reactions to that text and, in turn, to others. Why, for example, does one respond with particular intensity to a given scene in a given play? Perhaps particular images converge, rhythms change, sentences become shorter, adjectives suddenly follow nouns, phonetic patterns shift. The occurrence of all these events as well as the contrast of their use in the given scene with their use elsewhere in the play would be almost impossible for a given scholar to observe in a bearably finite period of time. At present, it is also impossible for the computer to observe them all because we, ourselves, do not know clearly how to describe some of the patterns, e.g., syntactic and phonetic, that we'd like the computer to observe. That is--and this is crucial--we have to pre-edit to enable the computer to do its work--and pre-editing is something we'd like to avoid. But the computer does hold out the hope for more comprehensive as well as more precise description, or modeling, of literary style. As students and scholars, we would like to see stylistic patterns revealed, in all their complexity, and relate the perception of such patterns to other of our responses. It is this kind of intellectual activity which we must have if we are to 'understand' the world in which we live. Indeed, it is our belief that it is, in part, toward this kind of understanding that education is directed.

* * *

Over the longer term we look to an emerging sense of matters of style as matters of considerable social consequence. The reason for this is not far to find: with established system relationships in which man is implicated (as with population, resources, and culture) being disrupted in increasingly consequential ways, the importance of the distinctive and significant human capacity for adequate modeling through symbolization grows rapidly in importance. Otherwise, without such modeling, the chances for re-establishing satisfactory systemic relationships would be greatly reduced. And as such symbolic modeling grows in importance we will come to see that we must effectively analyze human symbol systems (language, in a generic sense) themselves, and their dynamics, with great precision. We must do so in order to know well enough the properties--including the stylistic properties--of languages to match them (through redesign, when indicated) with sufficient effectiveness and efficiency to the properties of their central nervous system processors and to the properties of the dimensions of realities we may think it most urgent to model.

To repeat, the analysis of symbol systems and their dynamics is a central need of the present day; the distinctive and significant powers of man are a function of his ability through symbolization to adequately model aspects of himself in his environment, and as traditional system relationships are increasingly disrupted the social urgency of accurately understanding the properties of the various types of languages in which such modelling is done in order to use them more effectively assumes even greater salience. The computer is our one present hope as the instrumentation requisite for a comprehensive understanding of languages or symbol systems; and it is a widely held basic assumption that a computer or technological configuration including some form of computer will serve as critically important instrumentation in comprehensive research on languages.

There is no major problem faced by men for which the information available, the symbolic form in which that information is available, and applicable communication technology are not critically important conditions for any solution to it. Further, the very sense of there being a problem and the character of that problem are themselves functions of properties of the language modelings of those parts of the environmental domain in which the problem is symbolically 'found'. In some cases the mode and character of the symbolism which is thus problem-definitional may itself prove to be at least as much a part of the problem as of its solution. We say, then, that symbol manipulation is a denominator common to all of man's fractional efforts to cope with his environment; and, further, that understanding symbol systems with the greater thorougness needed to markedly improve our ability to utilize knowledge in the solution of our increasingly numerous problems is itself a matter of high urgency.

Both pragmatic and theoretical considerations increasingly impress upon us the importance of comprehending more adequately the ways in which languages/symbol systems affect the direction and control of the numerous, variegated, and interactive systems (social, natural, and technological) of which we are a part. Cybernetic and general systems research, as well as many middle-range and applied studies, unite in forcefully demonstrating the heavy weighting due symbolic communication as a variable-set in explaining the functioning of the human component in nature.

Man is notably differentiable from the rest of the world with respect to his capacity to model efficiently both himself and his environment. The entire progress of science and successes in man's effective relating to environment and environmental resources are manifestations of that capability. The tests of the importance and successful use of that capacity are in no insignificant way imposed on men from outside ourselves: our ability to survive as well as to achieve environments of high quality flows from our facility in adequately using symbolic inputs to model accurately and efficiently ourselves in our interactivity with all the impinging phases of our ecologies (human; natural;

artifactual; and mixed). Whether with people or with 'the forces of nature' the extent of our success in survival will be heavily conditioned by the adequacy of the internalized simulations of them we derive from the transformation of symbolism received as sensory in-put.

One of the major necessary conditions for a scientific understanding of symbol systems--including the computer hardware devices so important to it--is the precise measurement of languages--'natural', formal, and graphic languages. At present languages have not been even effectively parameterized. Further, as to the empirical phase of such research, we do not know in anything approaching a comprehensive way what dimensions of language have been discovered and measured; nor, in those instances where any given investigator is interested in normalizing with reference to some property of language, have we generally any compilation available for him of the corpuses of language which have been measured along the dimension of interest to him, nor the results of the measurings. In brief, we may say that our knowledge of language components of symbolic processes is both radically defective and unassembled, not to say not readily accessed. And all this is true despite the major efforts in modern linguistics in recent decades--which have concentrated heavily on formalisms for syntactic characterizations.

Many different languages or notational systems are in current use. These include the so-called 'natural' languages, programming languages, mathematical languages, kinesics (non-verbal human expression), two-dimensional languages used in computer graphics, languages for machine-to-machine communication, languages (as yet only sketchily formulated) used in photographic representation. Currently, it is not known which languages are best for which tasks. It is not known sufficiently how to gain a sense of the systemic properties of languages so that models of the languages can be built, tested, and redefined, because we do not know what the parameters or variables of the languages are and we do not understand their interrelationships. We do not know how to relate various types of languages to each other in any coherent, systematic way; we don't know what the trade-offs are, what the intermappings are. We do not understand clearly what one language supplies and another lacks; we don't have an understanding of the interfaces between and among types of languages.

It is possible that with a thorough-going understanding of the patterns formed in the linguistic encoding of information we could discover that certain styles were more efficiently processed by the brain than others, and that certain styles were better adapted to certain types of environmental modelings than others. We may surmise that as the desperate environmental plights toward which we seem to be so inexorably headed grow closer and closer our sense of the social urgency of understanding style for its interface significance as a component in the symbol systems man uses as mind seeks to comprehend matter will grow far greater.

If we do come to find that in the analyzing system of mind-symbols-reality certain styles in symbolizing make for markedly superior effectiveness, we may be certain that 'value judgements' will not be needed by us to choose to use those styles. Rather, in the very nature of the case that which is more effective will be used; what is more effective will be what is richer in survival value through superior power in enabling us to cope with our environments. As with science, in the large less adequate notational systems, and modelings in them, are not reverted to.

And in that comparison with the general history of symbolism in science--with the simpler of two models always expected to triumph when each of them deals equally well with the same domain--we find a clue to what may be the neural basis for such styles as will be found superior to others. Those styles which may be expected to persist when others disappear will be those in which the patterns are as complex as necessary in characteristics needed for taking into account those features and relationships which are our response to the realities we would relate to, and in which the patterns include no more redundancies than

make for optimization in the processing efficiency of human minds.

Such styles won't have to be chosen. In the systemic situations in which they'll be humanly employed they will have been, in a manner of speaking, automatically system-selected.

REFERENCES

Carroll, John B. 1969. Vectors of prose style. (in Dolezel and Bailey, 147-55).

Chomsky, Noam. 1965. Current issues in linguistic theory. The Hague: Mouton.

Davis, Robert G. 1970. Rimbaud and Stavrogin in the Harvard Yard. The New York Times Book Review. June 28. 2,38

Dolezel, Lubomir and Richard W. Bailey, ed. 1969. Statistics and style. New York.

Frye, Northrop. 1957. Anatomy of criticism: four essays. Princeton.

Kondratov, A. M. 1969. Information theory and poetics: the entropy of Russian speech rhythm. (in Dolezel and Bailey, 113-21).

Kroeber, Karl. 1969. Perils of quantification: the exemplary case of Jane Austen's Emma. (in Dolezel and Bailey, 197-213).

Lanier, Sidney. 1891. The science of English verse. New York.

Leed, Jacob, ed. 1966. The computer and literary style. Kent, Ohio.

Miles, Josephine. 1967. Style and proportion: the language of prose and poetry. Boston.

------- and Hanan C. Selvin. 1966. A factor analysis of vocabulary of poetry in the Seventeenth Century. (in Leed, 116-27).

Milic, Louis T. 1967. A quantitative approach to the style of Jonathan Swift. Studies in English Literature, vol. 23. The Hague.

Pierce, John R. 1965. Symbols, signals and noise: the nature and process of communication. New York.

Sedelow, Sally Yeates. 1970. The computer in the humanities and fine arts. Computing Surveys. 2.89-110.

------- and Walter A. Sedelow, Jr. 1967. Stylistic analysis. Automated language processing: the state of the art, ed. by Borko, 181-213. New York.

-------. 1969. Categories and procedures for content analysis in the humanities. The analysis of communication content, ed. by Gerbner, et al., 487-99.

Singh, Jagjit. 1966. Great ideas in information theory, language and cybernetics. New York.