FORMALIZED HISTORIOGRAPHY,
The Structure of Scientific and Literary Texts:
Some Issues Posed by Computational Methodology

Walter A. Sedelow, Jr.
and
Sally Yeates Sedelow

Affiliations: Division of Mathematical and Computer Sciences, National Science Foundation, Washington, D.C. 20550
and
The University of Kansas at Lawrence, Kansas 66045

Working paper for:

The International Symposium on Quantitative Methods in the History of Science

University of California/Berkeley
Office for History of Science & Technology

25-27 August 1976

"FORMALIZED HISTORIOGRAPHY,
The Structure of Scientific and Literary Texts:
Some Issues Posed by Computational Methodology"

Walter A. Sedelow, Jr.
and
Sally Yeates Sedelow

I.

Clearly, quantification can be useful for historians of science. Of course there may be illusions as to the precision of the quantification, just as there may be illusions as to implications of the quantification. And quantification certainly appears to have, both for better and for worse, a favorable aura deriving from not always well construed associations with precision and scientificity.

The broadest significance of what we have to say here is that, while (physical science) quantification subsumed and succeeded (positivistic) precision as at the cultural center of science, quantification itself partly with the help both of the computer and of computer science has been subsumed and succeeded at the core of science process by formalization. Similarly, prediction subsumed and succeeded description, and generation and control (formation and transformation), prediction as the most central proof of scientific knowledge.

We further argue that while computers are sometimes mistakenly seen as numerical precision machines, strictly speaking that is an incident--albeit a very important incident--of their essential property, which is logical and formal rather than numeric.

Since a general purpose digital computer can realize any properly formalized processes, its scope includes both processes involving quantification and others which are non-numeric. Hence the computer can in principle realize whatever we call "public science" and, necessarily then, its importance to the history of science is co-extensive with the entire overt scientific process, rather than only with its quantitative and quantifiable components.

Whether in the study of primary or secondary sources in the history of science, while quantification and subsequent statistical manipulation are critical and often machine-facilitated, non-numeric processes are also critical and, ironically, may in practice come to be even more machine dependent. Hence, formalization and computation are more appropriate than quantification among general aspirations in practicing the history of science.

What can computer-based systems do for the history of science now? What have we reason to believe they may be able to do in the near term? What have we reason for thinking they're not going to be able to do? How close could they come to emulating a historian of science? How might they change the character of the history of science, and hence, the nature of the historian of science? Those are the questions we seek to pose and suggest possible answers to.

In contradistinction to quantitative history (quantitative historiography), here we are primarily addressing ourselves to a 'second order' phenomenon: computational historiography

or, better, formalized historiography, wherein historiography stands not for the writing of history (as it does in quantitative historiography) but, by contrast, for the rigorous, mathematically-formalized analysis of historical writing.[1] It may be argued, of course, that the only non-illusory meaning of history is, in an extended sense, historiography (historical writing, as well as historical lecturing, etc).

In a sense we may say that there can be various types of artificial intelligence, or intelligent systems. In addition to the most generalized AI, we could also speak of Artificial Scientific Intelligence, with reference to, say, the work of Lederberg and Feigenbaum (e.g., the DENDRAL system) at Stanford[2] or, more loosely, the aspirations of the AIM (Artificial Intelligence in Medicine)/SUMEX group, at Rutgers and elsewhere.[3] Here, in this present essay, we address the more specialized functions appropriate to an Artificial Science-Historical Intelligence.

In this discussion we will try consistently to maintain a distinction between quantification and computation--with computation as the more general category of formal processes of which numeric processes are but one type. Quantified, numerical results are calculated by the machine through the utilization of its logical structure. There is, of course, a non-physical parallel of sorts--in that we treat mathematics as a sub-set of logic, rather than the reverse.

Since, unfortunately, computation tends to have certain stereotypic connotations heavily reinforced by association with calculation, it is perhaps not as easy as one might like to be certain that the generic logical (contra numerical) meaning of computation is held clearly and distinctly in mind. In part as a way to cope with that difficulty, we will here try to stabilize use of the terms formal, formality, formalize, formalism, etc. and the associated concept of algorization in discussions of computation, on the assumption that the 'formal' lexicon is less polluted by extraneous and inappropriate associations and assumptions about numeric algorithms. It also bears remembering that formality in the mathematical sciences entails meeting much more explicit, strict, and onerous criteria than the very different formalism of the humanistic disciplines.[4]

Quantitative historiography is sometimes employed to mean quantitative history, or at least the writing of quantitative history ("cliometrics"). We have used computational historiography rather differently. In computational historiography, historiography means

not the act of writing history, but the analysis (and history) of historical writing. (Elsewhere Walter Sedelow has sought to make clear that the distinction between history and historiography is fundamentally flawed, in implying an existence for history apart from, in an extended sense, historiography-as-historical-writing.)[5]

Let it also be remembered that quantification is not an end in itself, and that unless it is clearly understood as means there is the danger--possibly the likelihood--of its deterioration into modern numerology. Counting can become a compulsive activity, and when it does lose its grounding in functional rationality there is always the possibility that, as with the Pythagoreans of an earlier day, it will degenerate into a sort of mystical religiosity. Sometimes we see that these days, and label it a phase not of science, but of scientism. Of course we do need to remember, as Seneca intimated, how one person's madness may be another's genius, and that irrespective of the dignity of its 'motivation' sheer counting may contribute to interesting results. It may make possible the detection of a hitherto unperceived empirical regularity, which in turn could be a critical component in filling out a useful theoretical model.

Our own research has for many years included studies in the uses of computers as complements to intelligent human behavior. One of our orientations has been that 'intelligent' systems should be designed to augment distinctively human intellectual strengths,

to provide a mechanical means of doing what unaided humans need or want to do but can't do, or can't do easily or comfortably or efficiently. Thus, intelligent systems are not necessarily human-emulation systems; and thus they are in character more extensive, as well as otherwise differentiable, from so-called 'Artificial Intelligence' systems. We say so-called in part because, contrary to claims frequently made for them, AI systems generally do not in any significant sense simulate human cognitive processes. Rather, when they are said to be simulating human thought, usually at best they are simulating human symbolic behavior outputs by means of a transformation function operating on symbolic in-puts such as might be presented to a person or people.

In this present essay, in contradistinction to our own research efforts, we undertake for purposes of exposition and conceptual clarification to compare some of what can be done by machine with some of what an historian of science does. That comparison should not, then, be taken to mean that our own primary researches --some of which are alluded to in passing--are governed by an aspiration of replicating human CNS (central nervous system) behavior. It can, though, be illuminating for seeing just what computer-based systems can do to compare them with human behavior in a complex professional role--even when the systems were not designed in the first instance with human emulation in view.

III.

As you are well aware, one of the most powerful criteria for quality in scientific work is generality. As an aspiration for rigor in the history of science the more general form of the concept of quantification is also preferable: that is, the concept of formality is preferable in guiding our aspirations for rigor, rather than the concept of quantification alone. One might well argue that valuable though the rigor implied by quantification may be in establishing precision in our work, the concept of formality is the more appropriate to studies in the history of science, not only because it encompasses both the procedures through which quantification is achieved and many other advantageous procedures as well, but also--and more importantly--because it calls our aspirational attention to the central property of rigor in science and scientific scholarship. That most critical condition of rigor is expression in formally stipulated algorithms of the processes by means of which precise statements may be generated, whether that precision takes a quantitative or a logical form. In principle when we have a formal procedure, expressed in a formal language, we have as fully as possible a ruleful use of symbols and of whatever externally to themselves those symbols are utilized to control.

We are suggesting in effect that in seeking quantitative precision in the history of science there is, as it were, an implicit quest for scientificity for the historical scholarship itself. As you are well aware the goal of a scientific history is not new, but

the 'location' (meaning) of the goal has moved from time to time and author to author. Characteristically exponents of a scientific history, whatever its meaning, have been exclusivists, as have been those who have endorsed its alternatives--except, perhaps, in the extreme case of literary history, whose exponents seem generally to have regarded it as only one option, perhaps often assuming that the practice of their own métier implied others engaged in more scholarly historical studies. Be that as it may, here we do not espouse in an exclusivist spirit the goal of quantification for the history of science, nor even the, in our view, superior goal of formality. Contrariwise, were we engaged in a comprehensive assessment of methodological desiderata, we would strongly urge the great importance not only of intuition, imagination, and other traditionally humanistic facets of historical study, but even of a kind of poetical history, as well; we would also stress the claims of the canons of assiduity in the use of primary sources.

But insofar as quantification is normatively orienting we wish to urge three considerations: the one, (a) that quantifying activities be seen as proper subsets of efforts for formal rigor; another, (b) that in the use of the computer in the analysis of both sources and of historical writing, quantification is but one part, and that not the largest, in the more formal analysis of such material (throughout this essay, we mean formal not in the weaker sense in which it is employed in literary scholarship but in the stronger sense in which it is used in the mathematical sciences);

thirdly, (c) that in the movement--however slow it may be--toward including a classically mathematical formality within the history of science one is, in the most quintessential sense of a science, moving toward making that much of the history of science itself truly scientific.

By way of a modest elaboration as to the force of the concept formalism, it might be said that a fully formal definition of a process (e.g., any procedure for studying sources or writings in the history of science) would be such that the process was then in practice implementable on a computer and associated equipment. That is a stern test. It means, for example, that if one had a formally adequate scientific text reading machine, each and every elemental perception of a characteristic, property, meaning, or significance that a human reader would want to be able to 'find' in studying a scientific text would have to be found through the use of explicit, 'machinable' procedures if that specific detail were to be counted as successfully coped with in a formal way. A critical condition is that nothing can be supplied by human intervention to such a computer-based intelligent system once it is in use. That's the rub: the formally adequate system must be fully explicit and self-contained. And, obviously, if such a system is to be useful it must be very general, very wide in its range of applicability--e.g., it must respond fully to a wide range of texts. At present there is, of course, no such system. But there are systems which will do some tasks of reading more adequately than

even a cultivated, specialized human reader. And we can now descry the way to doing many more reading tasks. The scientific knowledge on which to base a science-history writing machine--which would automatically provide the transformation functions necessary to converting input sources (automatically read) into output written history (computer-system produced)--is still embryonic.

The structure of the remainder of this paper will be, first, to sketch out some of the current capability, the state of the art, for the formal performance of text-analytic tasks; then, in a sort of test-case, contrast that with some of what we can do in a moderately sophisticated traditional mode; next, to underline the contrasts between the two approaches and to say something about what may be ahead in computational methodology; and, finally, to offer some appraisal of possible impact on the history of science of forthcoming computer-based formal methods.

If, then, we treat quantification in doing the history of science as special cases of formalization in doing the history of science; and if we hold to a very strong test as to what shall constitute formal adequacy here, such that the formal system not only has to meet tests for logical rigor but also has to be <u>realized operationally</u>; then we can go on to ask how close we are to being able for the history of science to provide a formal machine to read both primary and secondary sources. And a still more ambitious requirement would be also to specify the transformational algorithms

for converting machine-read/-analyzed primary source historical data input into output in the form of historical statements.

In answering this question of what we can now do algorithmically in reading and analyzing historical material, we shall argue that the syntactic analysis has been done comparatively more satisfactorily than other facets of reading (such as in history of science material), that semantic and semantic-syntactic reading are less well developed but nonetheless now have some successes to their credit, and that through our own current thesaural research we seem to have in prospect a way of achieving formal capability for reading/apprehending connotation. To make clear what a machine would have to be able to do by programmed rule if it were to cope not only with the syntax and the denotation (or paraphrasable sense content) of history of science material but also with its connotation, we will then examine in a traditional, non-computer-based, non-formal, but hopefully insightful way part of a chapter, relating to the history of science and technology, out of a volume in "The Rise of Modern Europe"[6] series, and allude to another case study, of Hume's History of England.[7]

Since this overview of what currently can be done in the realm of formal, computer-based approaches to the analysis of natural language (e.g., histories of science and their sources) must be brief, it is also eclectic. Statistical methods will not be covered because our own research emphasis lies elsewhere, but that omission should not in any way suggest that we regard them as unimportant. The general approach here will be, first, to discuss three overlapping areas frequently used in the definition of computer-based natural language projects--syntactic parsing, semantic analysis, and content analytic approaches. Secondly, we will try to 'map' from these areas to categories frequently used in contemporary work on continuous-discourse analysis--categories such as span analysis, anaphora, formal ellipsis, role analysis, and others. Notice that these are formal systems for use in reading natural language, that is, proceeding through a text in the predominantly linear fashion which an historian of science would employ in reading; there is an implied contrast with the alternative of a statistical characterization of the properties of a text taken synchronically as a population of verbal objects.

Syntactic parsing receives first mention simply because it is a procedure which early received extensive attention in the context of computer-based-systems. More specifically, the context was provided by machine translation, and the failure of the combination of syntactic analysis and word-for-word dictionary look up to provide adequate machine translation--a failure that surely

could have been predicted--cast a pall over computer-based natural language processing systems which has only recently begun to dissipate, despite a sufficiency of success such that some translating systems are today in heavy use. For extensive surveys of the state of the art as to this research, articles by Grishman[8] and Benson[9] are recommended.

Historians should note first that no "general-purpose" parser yet exists for any natural language. That is, it is not possible to select, for example, an historical text or section of text at random, present it to a computer-based parsing system and expect to receive a satisfactory syntactic description of the text. It is also quite likely that more than one parsing for a given language string, usually a sentence, will be provided. When a human being analyzes a text, he usually picks one analysis as best; in fact in consciousness he often produces only one analysis. A human being performs restrictively because he is aware of the context in which the language string occurs both in terms of semantic categories and of the section of text in which the particular string is embedded. Computer-based parsers have not had such extra-sentential awareness, although that situation is beginning to change, as will be evident in examples later in this paper of computer-based systems handling simple anaphoric reference and ellipsis.[10]

Granted the disappointing fact that there is no general purpose parser, still it is the case that parsers do exist which have performed well on restricted bodies of text and which may be

reasonably amenable to 'tuning' so as to be suitable for texts from other universes of discourse, such as a body of historical material. An example of such a parser is William Woods' Augmented Trasition Network parser which has received its fullest description within reports on the analysis of connected discourse concerning samples of rock taken from the moon.[11] This parser embodies notation from graph theory and ideas from automata theory and is susceptible of formal description.

Remarks concerning the capability of Woods' grammar as well as others one might mention must be made with caution because, as Benson points out, "there are no commonly accepted measures of syntactic capbility."[12] Both Woods' parser and the parser devised by Terry Winograd for his natural language dialogue concerning a block world[13] combine syntactic analysis with some semantic analysis. Again, as Benson points out, although these two systems

> make explicit use of parsing theory and implicit use of tree automaton concepts [for aspects of syntactic and of semantic processing], there is no theory which can be used to compare the differences in the way that the parsing and transformations performed by the tree automata are interwoven in these designs. What is needed is an algebraic structure theory for such devices. Such a theory, if successful, would suggest the appropriate decomposition of the total syntactic-semantic component into subroutines for implementation.[14]

Despite the practical and theoretical successes of syntactic parsers, there are several practical problems for which currently no theoretical solutions exist. Two such problem areas

pinpointed by Benson are discontinuous constituents and conjunctions. His example of a discontinuous constituent is "the detached particle in verbs such as drop...off in sentences like John dropped Mary off."[15] Computer solutions to the treatment of such discontinuous constituents as well as of conjunctions, thus far are ad hoc. Benson notes that although ad hoc solutions exist for these and other problems, there are no measures of their "quality, speed or effectiveness."[16] Benson goes on to indicate that in his view the most serious problem facing theorists is the non-existence of a "theoretical study of extensibility." By extensibility, Benson means "the ability to define new words and phrases at will, and to establish new meanings for existing words and phrases, these meanings to apply in particular context."[17]

The above comments by Benson concerning extensibility clearly broadly overlap semantics; so here we will make a transition to some brief comment on current capacities for semantic analysis.

Extensibility, which has just been mentioned, has been incorporated in several approaches to languages, the best known for natural language being the REL (Rapidly Extensible Language) developed by Fred Thompson, Bozena Dostert, and associates at The California Institute of Technology.[18] The notion of Thompson's REL is to permit the user to extend the processing capability--both syntactic and semantic--of a computer-based language processing system so as to meet the requirements of a particular body of natural

language text. This is an intuitively attractive approach to more adequate solution (for individual historical specialties, for example) of natural language processing requirements. But, as Benson points out, there should be theory which would "suggest data representation techniques and the appropriate balance between extensibility and a priori syntactic and semantic structures."[19] Extensibility is a research focus within which significant gains could be made; and it should be noted that Thompson's REL is being used by some social scientists today.

The use of various symbolic logics for representation of natural language sentences, including their presumed semantic import, is another approach which has had its attractions for implementers of computer-based systems. All of you who have studied such logics will remember textbook efforts to make the logics intuitively accessible to the student through the use of natural language statements; you will also remember the failures of 'fit,' (e.g., truth value assignments for implication in the propositional calculus which produced, at best, very tortuous natural language analogues), suggesting that the logics weren't entirely suited to natural languages. Nonetheless, logics have the undeniable virtue of being formal systems, and combinations of propositional and predicate calculi are frequently used for the representation of very restricted sets of 'knowledge' in computer-based artificial intelligence programs. For example, programs that compute optimal routings between two points or optimal strategies in a chess game may well use 'theorem-

proving' dependent upon 'logical' statements. An early, interesting effort to work out a logical representation of subsets of natural language is described in "Automatic English-To-Logic Translation in a Simplified Model" by Herbert Bohnert and Paul Backer.[20]

Recently, publication of notes and papers on Montague Grammars has resulted in rather intense effort by linguists, logicians, and computer scientists to develop this logic-based formal system for natural language and put it into computer-accessible form for testing and possible extension. Such a grammar provides a close and formal relation between syntax and semantics and has perhaps been most comprehensively discussed in a collection of papers edited by Barbara Partee.[21] Joyce Friedman, a logician and computer scientist at the University of Michigan, is currently involved in an effort to get parts of the grammar into computer-accessible form.

Some other current approaches to semantics within the context of computer-based systems tend largely to ignore any kinds of syntactic categories, choosing instead to focus upon the semantics of a particular universe of discourse.[22] Efforts have been made to define 'primitives' which are not restricted to any universe of discourse; but, even given such 'semantic function classes,' the meaningful detail specific to any given universe of discourse is considerable. Thus, current systems which deal with coherent discourse in a comprehensive, 'meaningful' way, are very restricted indeed and certainly could not cope with even the semantic facets

of any text selected at random in which an historian of science might be interested.

In an effort to cope with, inter alia, some 'frame of reference' aspects of semantics, we ourselves have been working toward formal description of a general-purpose thesaurus. This work is seen to be significant for content analysis and it is to that area that we next turn our attention.

Content analysis was a technique well known to social scientists prior to the use of computers and with the advent of computers it has received strong impetus.[23] Most, if not all, content analytic procedures have depended upon word counts which have then been used in statistical descriptions concerned with clustering, co-occurence, and so on.[24] In information retrieval tasks lists of words specially prepared for a particular topic or area are often called, somewhat loosely, 'thesauri.' Our research has led in a different direction, putting into computer-accessible form a general purpose thesaurus--Roget's International Thesaurus[25]--and then performing studies on that thesaurus with a view toward a general-purpose solution to some aspects of the problems of semantics.[26]

One of our concerns has been the notion of semantic distance. For example, if some of the current computer systems entailing research on semantics were dealing with the word "ship," it would be necessary to develop a conceptual framework indicating the import or imports of "ship" within the discourse being analyzed. Such efforts--which are being made in current systems--are responsible

for the very limited scope of those systems and suggest an ultimate requirement either for 10,000 dedicated graduate students or, better, for some other approach, if we are to have much semantic competence within the foreseeable future. Our hope is to use a thesaurus to gain a 'fix' on the context in which the ambiguous "ship" is being used in a particular segment of text. An experiment using Roget's Thesaurus to determine its usefulness for a related, practical computer processing problem--deciding when a given sequence of letters is a word prefix and when it isn't--has strongly suggested that a thesaurus might well be a useful guide to semantic distance.[27] If Roget's Thesaurus were used for this purpose it would require modification but it is almost certainly the case that such modification would be considerably more expeditious than building almost from scratch other interpretations of words' semantic relationships.

It is toward the goal of meaningful statements concerning semantic distance that a graduate student associate who is a mathematician, Robert Bryan, has been working on the graph theory necessary for the formal connectivity description of a thesaurus, with Roget's International Thesaurus serving as an instantiation. As Benson points out when discussing several current computer-based semantic systems, "There are an enormous number of possible structures for language, ranging from very shallow syntactic structure to very deep semantic structure"[28]; he calls attention to the strong requirement for some theory which would permit description of these structures, noting that "until such time as

the mathematical theory of formal languages develops to include these 'natural' language understanding systems, they remain ad-hoc."[29] Benson goes on to say that "for processes as complex as these, neither a verbal description nor a computer program listing suffices. There must be an intermediary of mathematical theory."[30] It is our hope to contribute to such a theory with reference to semantic distance, a concept which can then be incorporated into well understood computer-based procedures for language processing.

Until such time as we have a better theoretical grip on the area of semantics, it will be possible, as it is now, to use thesauri or analogous word lists to look for clusterings of related words in text. But it should be understood that although such work may well be pragmatically useful (and we have demonstrated the usefulness of such results for looking at a Shakespearean play)[31] the formal viability of such results will be open to question and argument. That is, for example, any word list or thesaurus may well be biased--and one of our goals in the study of Roget's International Thesaurus, obviously, is to try to discern just such structural instances of bias.

In summary, syntactic analysis of natural language strings is sufficiently well understood so as to permit computer-based systems to provide satisfactory language analysis and response generation within a universe of discourse such as that concerning the rocks found on the moon. It is also the case that syntactic analyzers are much more well understood formally than are semantic analyzers.

Semantic analysis is still in a preliminary state. For practical work on a large corpus of historical text, at present the most promising approach would be some version of content analytic procedures. In time it seems likely that resources useful for content analytic approaches may well be combined with programs dealing in a procedural way with semantics to provide a much improved approach to semantic analysis.

Turning now to the analysis of sequential, connected discourse, and to the categories employed for that purpose, what can be done in a rigorous, formal way?

Some current computer-based language processing systems can deal with restricted aspects of anaphora and of ellipsis. Anaphoric reference, as you well remember, means the use of a word or a group of words to refer back to words used earlier. A frequently cited anaphoric reference is a pronoun which points back to a proper noun used earlier in a sentence or in an earlier sentence. Computer programs are available which now deal with such reference. As an example of a program which deals both with simple anaphoric reference and with simple ellipsis we cite a few sentences from a paper recently published at the Stanford Research Institute:

> "What is the length of the surface displacement of the Lafayette? . . . What is its draft? and "What is the length of the Lafayette? . . . The Ethan Allen?" In this sequence of four sentences there is an example of anaphoric reference through the use of "its" in the second sentence and of elliptical anaphoric reference in the fourth sentence which, in its entirety, would read "What is the length of the Ethan Allen?"[32]

A very simple, and quite illuminating approach to aspects of discourse analysis described by Joseph Grimes in The Thread of Discourse[33] is span analysis. One type of span is an identification span, consisting of a series of identifications of the same participant in a narrative--a series in which no identification is stronger than the one before it. Strength of identification is defined as a ranking that goes from proper names, to explicit descriptives like "the last speaker at the symposium," to common nouns like "the speaker," to nouns used generically like "the fellow," to pronouns like "him," to reference without identification.[34] When a shift occurs from a weak form of identification to a stronger form, the current span is terminated and the new one begins. The spans can be shown graphically simply by taking a piece of paper, listing vertically from 1 to N the number of clauses, or phrases, and drawing vertical lines to shown the length of the spans. Obviously, computer systems could print out such spans or display them graphically on a display terminal.

Spans can be used for many characteristics of discourse, such as setting, time, and place, and can reveal patterns which graphically illustrate structures in the work being studied. In short, any elements which are substitutable or which fall within a particular set (such as terms referring to setting), could be studied in terms of spans.

Since students of history of science and of scientific discourse are concerned with the notion of parsimony, it might be

useful, for example, to utilize spans for the study of patterns of usage of substitutable elements, such as word synonyms, etc. Presumably, elaborate substitution would be less parsimonious than frequent repetition of the same words and phrases. It is possible to see how it might currently be possible to use a thesaurus within the context of a computer-based system to discern spans of synonyms or of words very closely related semantically. At least, it would be possible to experiment with such notions on the same basis that we have experimented with Roget's International Thesaurus in the area of prefixation.

Another form of analysis devised by Grimes and his students is role analysis.[35] This approach depends upon the roles taken by participants in a narrative. Roles are described by such terms as agent, experiencer, instrument, and other analogous categories drawn from case grammars. The focus is upon the shifting of roles among the participants in a discourse. The perception of shift is based upon relative rankings among cases such as agent, experiencer, and so on. In this system agent is the highest ranking role and a case called essive, which is rather analogous to the existential state, is the lowest ranking. In the folktales examined by Grimes and his students, patterns of paragraphing are clearly related to shifts in the roles of participants.

It is clear that in the historical study of scientific texts, it would be desirable to have the ability rigorously to discern entities filling such roles as agent, experiencer, and so on. A

computer-based system to identify such patterns in unrestricted text in unrestricted universes of discourse would appear to be far in the future. It might be possible, though, in the case of a particular text in a restricted universe of discourse, to identify words and phrases within the discourse which could then be analyzed for syntactic role and for sufficient aspects of semantic context so as to be recognizable as filling roles such as agent or experiencer. Even modest success here would be an impressive accomplishment, given the current state of knowledge about semantics.

Word connotations comprise another area relevant to discourse analysis. Work on thesauri clearly is applicable here. The semantic net in the thesaurus specific to a given word would surely provide some useful indication of connotation. If one were searching for irony or sarcasm, it might be possible to compare a semantic net or nets from the thesaurus with clusters of words actually appearing in the text. This latter possibility is only that; much research would necessary to establish viable approaches. The search for straightforward connotation would look to be a more tractable problem, but decisions as to search strategies would necessitate a much more thorough understanding of the structure of a general-purpose thesaurus than we now have.

Perception of the conceptual organization of a text is, of course, a prime component of discourse analysis. In The Thread of Discourse, Grimes tries to get at semantic representation of discourse by applying notions from generative semantics. Neither the

theory of generative semantics nor its application are sufficiently worked out to be of use currently for computer-based systems. The most successful approaches at present to the conceptual structure of long natural language strings use various tools from content analysis. Thus, we have used a set of programs we call VIA (Verbally-Indexed Associations) to look for semantically-related words in a text and then we employed some very simple routines called MAPTEXT to display the occurrence patterns.[36] These patterns, of course, comprise an important component of conceptual structure. John Smith of the Pennsylvania State University's Department of English has used some of the VIA programs and other programs of his own divising employing principal component analysis to display patterns of images in a literary text.[37] Many of these procedures entail human intervention at some point. In our programs, that human intervention point is the selection of a word occurrence frequency threshold. In Smith's programs, that point was the initial identification of the image-bearing words and phrases.

Having now shown the current scope and limitations of formalized--if not always fully formal--systems which could be applied to reading English science-historical text by machine, the succeeding section of this essay attempts to still further test the adequacy of extant mechanical analysis by exemplifying aspects of a reading process for the history of science which, at least currently, would be especially difficult to even attempt to cope with by an electronic automaton.

In order to make clearer and more explicit some of the differences between reading as an historian of science might do it and as it could be done at present by a computer system, we have here a selection of passages from the third chapter of C.J.H. Hayes, A Generation of Materialism;[38] accompanied by commentary we have supplied. That commentary, and any inferences as to what it may imply in the way of human reader analysis of text, can then be a basis for contrast with computer-based reading. The discussion of computer-based reading procedures in the previous section of this essay should provide a sound foundation on which to build knowledge of what sorts of processes are at present out-of-scope for intelligent systems. As out-of-scope we stress here the attitudinal factors and associated connotation which text may invoke in a human reader. Insofar as--contrary to the emphasis in our own research--an attempt is made to replicate human symbolic output as a function of input, then, it is evident one of the major deficits in such emulation is in parsing for the non-referential, non-instrumental dimensions of text as they impact readers differentially. Those differentials are products of what the reader supplies to the reading experience--both by way of prior experiential input and by way of general and specific genetic factors.

A full capability for replicating human matches of outputs to inputs would, then, imply a system simulation of the human reading situation. We do have approximations to various constituents of that reading process; in addition to the syntactic, semantic, and other

sub-processes discussed earlier, we could also mobilize in a fairly organized way in a programmed system some of the knowledge we have as to attitude formation variables as they are made manifest in the ways in which people differentially generate language acts in response to verbal stimuli. There are also process models, as in the work of Robert Abelson,[39] which provide us with a basis for simulating the effects of personal dynamics in producing the attitudinal components of verbal outputs in response to verbal inputs. The extensive research on what makes reading difficult is also relevant in understanding how, in an extended sense, a text will be parsed.

Here follows an excerpting from Hayes' third chapter, along with 'reading' commentary on facets of the passages quoted:

> Page one hundred six: "Lister's discovery of antisepsis had been anticipated by the Hungarian Ignaz Semmelweiss whose achievement, however, had been neglected by the rest of Europe much as Mendel's was."

"Neglected" presumably implies willful action and even culpability, with attendant overtones of moral wrong. The denotative as well as connotative inappropriateness of that word in this sentence can be seen when one notices the final, elliptical phrase, "...much as Mendel's was"--"neglected" understood. But elsewhere in this chapter there is also an assertion which points up the dubious propriety, or at least the inconsistency, of these connotations of moral wrong: on page one hundred twelve it is noted that Mendel's important discovery "...was buried away for thirty years in dust-gathering tomes of a local scientific society." Even within Hayes' constructed frame of reference, is "the rest of Europe" then guilty

of "neglect" in not learning of it?

> Page one hundred seven: "Even Liberal regimes did not cavil about violating the sacred precepts of laissez faire and invoking the most stringent police powers in the cause of public health."

If one reads this sentence with the first word, "even," omitted, then one notices the difference in tone and, therefore, the force of "even." One could wonder whether Liberals spoke of laissez faire principles as "sacred" with a frequency sufficient historistically to warrant the adjective here; if they didn't, with the agency for the "sacred" left unindicated, perhaps we have to conclude that it is the chapter's author who is introducing the word, and in doing so introducing what amounts to a sneer at these laissez faire principles. (If there is any doubt about the sneer, read the sentence while omitting both "even" and "sacred.") This sentence also manifests, in, admittedly, a rather minor way, a successful employment of the have-your-cake-and-eat-it-too principle, as is apparent when one notices that the appropriateness of the "even" and "sacred" even as sneers (in suggesting something hardly credible in Liberals) has been vitiated by such remarks as those on page eighty-three introduced by, "Nor did the Liberals evince any squeamishness about invoking in behalf of popular education that very principle of compulsion which they were credited with abhorring." (It should be noted that Hayes didn't precisely say the Liberals abhorred that principle, only that "they were credited with abhorring" it.)

Page one hundred eight: "MECHANISTIC NATURAL SCIENCE"

Here we note the adjectives employed in three successive section headings beginning with this one. Considering the treatment in the text, for each of these adjectives--"mechanistic" with natural science, "deterministic" with biological science, and "physiological" with psychology--and there is, incidentally, an obvious inner harmony among them and with the rest of the chapter and book, each of these adjectives in effect proposes that within its respective discipline the approach suggested by the adjective was the only one taken.

> Page one hundred eight: "Such a conclusion seemed to be inescapably drawn from the then known facts..."

"Seemed" to whom? Notice that agency is undesignated, that this has particular effects here comparable to those previously indicated for other examples of this phenomenon, and that the passive voice has a role in disguising the existence and the tacit nature of the implications involved.

> Page one hundred nine: "Obviously the whole universe was constructed of the same material elements."

Agency undesignated again, and again the seductive passive voice in the background.

> Page one hundred ten: "...physical science...appeared.. quite secure and well-nigh complete. In the future little would remain to be done, it was imagined, beyond measuring physical constants..." "Further investigation, it was predicted, would prove this assumption--though the next generation of physicists learned with shock that it didn't."

These quotations are all drawn from the same paragraph. First, note that agency is undesignated. Then, note how the main verbs are equivocal: "physical science...appeared"; "little..to be done, it was imagined"; "investigation, it was predicted." And notice, finally, how that all helps to build high for a longer fall: "though the next generation of physicists learned with shock that it didn't."

> Page one hundred eleven: "Unfortunately for the certitudes of physical science, a delicate experiment of two Americans, Michelson and Morley..."

To feel the force of the introductory "Unfortunately... science" phrase, substitute for it, as the qualifying connective, the single word "However." (Taking the paragraph as a whole, no denotation is lost through effecting this substitution.)

> Page one hundred eleven: "Just as physical science inspired confidence in its mechanistic and materialistic assumptions...so biological science, by its promise..."

We have in the use here of "physical science" and "biological science" an instance of false reification, and it is a kind of variant on agency undersignated. By speaking of "physical science" and "biological science" as, in effect, things, the reader's mind presumably is carried away from the unique complexity of (the work of) individual, differentiated, physicists, chemists, and biologists, and the recognition of diversity among them. Instead we may be carried toward thinking as if there were no individuality, no differences in view and practice among them. The effect of this manner of expression is to condition the mind to be more receptive to the highly generalized, if not to the overly generalized.

> Page one hundred eleven: "In a period when incredible as it may appear, health was even more eagerly sought after than wealth, the novelties of biology naturally attracted more attention than the somewhat staid and prosaic course of physics."

What of the soundness of such an extraordinary generalization as that in Europe from 1871 through 1900 "health was even more eagerly sought after than wealth"? Clearly neither in behavioral data nor in public attitude data have we the basis for a well-grounded generalization on this topic. The "incredible as it may appear" has a sarcastic effect contingent upon one's accepting, perhaps in part from the context provided by this book itself, certain general views of the period in question. And there also seem to be overtones about human nature, or at least about the people of Europe from 1870 through 1900, in the suggestion that "the novelties of biology" were "naturally" more attractive of attention than "the somewhat staid and prosaic course of physics."

> Page one hundred twelve: "At the same time it was well known, at least to practical gardeners and farmers, that new varieties of plants and animals could originate in 'sports'..."

Since we may suppose that "practical gardeners and farmers" is not to gain its meaning through implied contrast with "impractical gardeners and farmers," presumably, then, the "practical" serves to discredit those whose interests in heredity were "theoretical," i.e., the"Darwinists," who are thus represented as not to have known what practical farmers knew all along.

> Page one hundred thirteen: "As we remarked in the first chapter, the distinctively Darwinian doctrine of natural selection attained a great vogue in the early '70's, partly because of its simplicity and seeming applicability to a wide range of human interests, and partly because of its concurrence with a high tide of industrial and military competition. The vogue remained throughout the era and gave continuing direction to a vast deal of inquiry, not only in biology but in psychology and the so-called social sciences."

It is noteworthy that there is no indication here that even among professional natural scientists there might have been some who were 'motivated' in their investigations primarily by a pure interest in biology and a quest for truth. Rather, the interest in Darwinian ideas is attributed to their "simplicity," "seeming applicability to a wide range of human interests," and "concurrence with a high tide of industrial and military competition," none of them motivations with, in context, the suggestion of highest seriousness. And the interest in Darwinian ideas is twice referred to as a "vogue"--not a flattering term, certainly--especially inasmuch as no explicit provision is made for less irrational factors.

One misses the differentiation, conventional by 1941, of the Social Darwinists from the more exclusively biological investigators. Perhaps the simple use of such a differentiation would have placed Darwinian concerns in a less trivial light. Note, too, the phrase "so-called social sciences"; one is tempted to suggest that they are "so-called" because they are so called, and to wonder whether this use of "so-called" leads to a confusion over whether methods, or conclusions, are the hallmarks of a science.

> Page one hundred thirteen: "...such a mass of confirmatory evidence as to leave no doubt in the mind of any well-informed person that all life was essentially one..."

The "as to leave...person" phrase which in another context can well be imagined as either neutral or favorable in connotation, here seems to take on coloring from its surroundings, and become somewhat sneering.

> Page one hundred thirteen: "...in total ignorance of Mendel and his work they [i.e., Darwin and his disciples] went gaily on their way, brushing aside the specialists in heredity as though they were mosquitoes, and blithely assuming that natural selection was the proved and adequate cause of evolution and the origin of species."

The question of the significance of ignorance of Mendel was discussed above. There would seem to be no need to demonstrate the existence of heavy sarcasm in "in total ignorance...they went gaily on their way, brushing aside...as though...mosquitoes, and blithely assuming..."

> Page one hundred fourteen: "This amusing family tree was presented quite seriously by Haeckel..."

The putative information that this schema was presented "quite seriously" might perhaps have been omitted, if "amusing" hadn't already been inserted. Here one gratuitous observation begot another; and also the "quite seriously" implies the historical fallacy of holding a past responsible for the knowledge available only at a future date.

> Page one hundred fourteen: "...Helmholtz, under the spell of Darwinism..."

A case of "perspective by incongruity": a special sense of the nature of Darwinism, and of Helmholtz, through likening Darwinism's effect to that of the superficially incongruous, i.e., a "spell."

> Page one hundred fourteen: "And Darwinian social scientists imagined even greater marvels."

A pejorative impact elsewhere achieved by linking "so-called" with social scientists may here be achieved through the use of "Darwinian." The rest of the sentence is sarcastic, and makes use of play on a previously undermined traditional contrast between the religious and the scientific.

> Page one hundred fourteen: "The vogue of Darwinism..."
>
> "Vogue."
>
> Page one hundred fourteen: "...this fed the belief that all the various activities of living organisms would presently be disclosed as mere modes of atomic motion and manifestations of mechanical or chemical energy."

The agency undesignated problems are here raised by the absence of any indication as to who held these beliefs. One should notice also how heavily pejorative "merely" is--to which word a measurement-by-omission test may be applied.

> Page one hundred fifteen: "Haeckel's dogmatic book of revelations..."

"Revelations" is sarcastic, as also the full phrase "book of revelations."

> Page one hundred fifteen: "Though direct evidence for this conclusion was unluckily lacking, it was widely accepted on faith, proving that even with scientists, or at any rate pseudo-scientists, faith may transcend knowledge."

In context "unluckily lacking" carries a note of sarcasm, and one wonders what criteria have been employed here to separate scientists from "pseudo-scientists." The comment on the faith of scientists as transcending their knowledge is so stated as to have about it an air of scoring against science; from the standpoint of recent decades, with non-rationalities in scientific process widely acknowledged, one wonders if that phrase is appropriate in its connotations. (The feat of having-your-cake-and-eating-it-too is also achieved here, when one notices that the tenor of the "this conclusion...was widely accepted on faith" statement is at considerable variance with another, on page ninety-eight, that "Certain fables of ancient Phrygia...in a modern Europe otherwise most critical of myths.")

> Page one hundred fifteen: "The rise of 'scientific' psychology..."

Author dissociation and disapprobation are felt in the single quotes around "'scientific.'"

> Page one hundred sixteen: "Its [i.e., scientific psychology's] spirit, if one may so denote a very material thing..."

This jab is, of course, based on a pretended confusion of one senseof the word "spirit" with another.

Page one hundred sixteen: "...Sigmund Freud was to tackle the whole problem of psychological abnormality, and his fame would outstrip Lombroso's."

The tone-through-omission effect in this sentence is detectable when one notices that Freud, being compared with Lombroso, is said to outstrip him in fame--and it is left at that, as if amount of fame were the sole significant measure of the one man against the other. Given stereotypic responses to Lombroso, to compare Freud with him in this way is certainly not advantageous to Freud.

When we examine what is currently possible in automatic response to text, and compare the outputs of those transformation functions with outputs by a human reader, say an historian of science, we see in the discrepancies (except for those where a reader is less adequate than the mechanical processes) work that some computer scientists, linguists, and other social scientists doing language research may wish to try to accomplish. When suggestions are made as to what may be promising means for rendering tractable these tasks, the suggestions can partake in differing measures of the pragmatic and the theoretical; insofar as possible we shall try to suggest lines of research which are theoretically based, in the interest of both cumulativeness, and especially, generality, with its attendant expectation of breadth in scope and applicability. Automated language-analysis systems which are only ad hoc offer little promise of economic utility, scientific value, or intellectual substance. Occasionally a lucky scientist may, of course, have the good fortune of inventing/discovering without guidance from a controlling model an idea and/or procedure which he, or someone else, can turn to very extensive good account in an exceptionally large population of cases. But our reading of the history and character of science intimates to us that such serendipitous good fortune is more likely to occur to the person who explicitly attends to the desirability of theory construction whilst also working on the experiemntal and the engineering possibilities in a research domain.

Where, then, do we now stand with reference both to theory and to practice in reading/analyzing text, such as could be of interest to the historian of science?

As you know, although our successes in parsing sentences syntactically by machine are more considerable than any other aspects of our capability for reading and understanding by automaton or electronic robot, there are still severe constraints on what of this sort can be done. It should be borne in mind, at the start, that any machine-based parsing will, at the present time, generally require that each word which is parsed be locatable in a machine-readable dictionary within which each word has attached to it one or more grammatical function designations. Further, as you know also, although there is increasing interest and successful work in the design of machine reading systems for languages other than English, it is still the case, especially if one excludes machine translation systems, that most of our accomplishment has been confined to the English language. Further, there is no general purpose parser even for English text. And such parsers as we have are not notable for their efficiency in generating the one most satisfactory syntactic reading of a sentence, even though some parsers are notably more satisfactory in this respect than others. Further still, the current range of discourse domains within which parsers function at all satisfactorily are really rather restricted, against the needs of an historian of science for breadth

of capability in a machine which would be of use in his work, much less against the criterion of a parser usable for any English encountered.

Nonetheless, as indicated elsewhere in this paper and as you can see much more fully from the literature which it cites, there are impressive accomplishments, and not only as to parsing which is syntactic. We won't attempt to recapitulate here what we have already said earlier on in this paper as to those accomplishments. But we do want to stress that insofar as capability for the historian of science is concerned, there are various procedures which can be machine-implemented now, the availability of which is forestalled only by their current relative cost.

It probably also deserves underscoring that such parsing as we have had to date has primarily been focussed on units of sentence scale or smaller. Only recently have we begun the serious work of attempting to, in a larger sense, parse extended natural language discourse of scope greater than single sentences.'Discourse analysis,' as this kind of supra-sentential work is called in the United States, is of course of far greater interest to the historian of science than attempts at exclusively sentence by sentence syntactical analysis. It is of course also true that there is now an interest in a generalization of the concept of parsing so as to carry it over to the structural properties of other media than those in which language is expressed, so that there is an interest in the parsing of videotapes and films and human costumes, for examples.

As to the semantic facets of such texts as the historian of science might be interested in, whether those texts are primary sources or secondary analyses, we can do somewhat less well at the level of the meaning of the individual sentence than we can as far as its syntactic properties are concerned; but with reference to extended discourse we can do somewhat better as to the meaning component than we can with the syntactic and rhetorical components. The approaches which are taken to getting at the conceptual structure of discourse may be divided into those which are semantic in character and those which are content analytic in character, with provision for mixed cases. If the semantic and the content analytic approaches describe the two major sets of techniques which we bring to the text, with reference to the texts themselves our ability to cope with meaning, by comparison with syntax, is also dual. One of the kinds of meanings which we can deal with has to do with the frame of reference component in the semantic portions of text, while the other kind of meaning is the operational significance of meaning content of text; and, again, there are, if we have so divided our study, mixed cases, where both frame of reference or universe of discourse kinds of meanings are involved at the same time that instrumental significance sorts of meanings are at stake.

Against criteria for good science the procedures which have been developed for semantic analysis of text are found wanting, in that they are grossly deficient in generality. The semantic understanding subsystems which we have are highly ad hoc and in each instance

confined to very specific kinds of knowledge. Since, as indicated earlier, those kinds of knowledge are of two types, the propaedeutic (which establishes the frame of reference function) and the instrumental, it may be worth mentioning that procedures for coping with frame of reference content in the history of science may prove to be especially useful with reference to older primary materials which are more 'theoretic' in an old fashioned sense of that term or, if empirical, embodying what we now perceive as a pseudo-empiricism. Insofar as the content of history of science materials is experimental in character or where there is a clear epistemic correlation between statements in the text and indicated behaviors by a scientist using the text, then the operations-oriented systems for responding to the meaning of text will seem the more useful. For example, it would seem as though the process oriented system of Winograd[40] would be useful with procedure-indicative components of scientific materials. One might think, for example, of the transferability of procedures utilized in assembling pumps. Some of the systems currently being utilized in unmanned space exploration, as in the Martian instance, presumably would also be cases in point.

As to content analytic systems, they are our current best choice for getting at the conceptual structure of strings of greater than single sentence length, as they are also our method of choice in trying to deal with the difficult business of connotation and attitudinal factors exemplified in the Hayes excerpts and commentary. But content analytic approaches are not satisfactory in giving us

ways of generating operationally correct meanings when the significance of a component of text is primarily instrumental. Our own research in conjunction with the VIA programs is relevant to attacks on the problems of connotation, while our thesaural research is very significant, we feel, for the effort at systematizing the elucidation of overall frame of reference for bodies of text.

It would seem that the application of further developments in thesaural-utilizing systems for establishing frame of reference and VIA-like systems for coping with conceptual structure within a frame of reference, when both are merged with operationally-oriented semantic schemes would give us the best presently visible systems for electronic robot responses to materials in the history of science.

Probably the contrasts between what we are now capable of doing by machine and what has been done traditionally in responding to passages out of the Hayes chapter do not require much heightening of emphasis. Nonetheless, it might be useful to point up here some of the limitations in our current repertory of intelligent system capability.

With reference to the first passage quoted, that apropos Lister and Semmelweiss, we are only now able to make a start at dealing with the simple problem of ellipsis that occurs in this passage. As to the larger connotative implications involved in presumed imputations of culpability, it would seem that any machine-based ability to deal with that component would have to be dependent on first mapping the text into a logical structure capable of revealing the contingent

relationships among the paraphrasable sense contents of individual sentences, as well as the more monumental task, in many instances, of having in machine readable form large corpora of filtered empirical data, such as, for example, the Encyclopedia Britannica. Walter Sedelow has dealt with an aspect of this problem in his discussion of a possible Bicentennial Electronic Encyclopedia for the United States.[41] At a more theoretical level issues relating to this question of how to render tractable the determination of these connotations may be found in early work of Ross Quillian.[42]

For understanding the passage from page one hundred seven of Hayes, in principle one could try to combine a parsing which was confined to the referential and instrumental meanings of a sentence with a carrying over of it for checking against its attitudinal significance within the framework of an organized body of contemporary psychological research, as in the General Inquirer programs developed under the leadership of Philip Stone.[43] Such an effort would be not only very complicated but it would be inherently vitiated by the essentially consensual validation of the taggings of terms done for the General Inquirer programs.

If there were in a parsing program a requirement that the agency in any sentence be flagged and its absence also noted, then one would have at least the empirical base for attacking the problem of agency undesignated as, for example, with the passage on page one hundred eight of Hayes. But its role, as well as the role of the passive voice--which also could have been flagged--is something else

again. The significance of the absence of specification of an agent is, in this text at least, sufficiently great that one might wish to have the absence of agency clearly signalled.

With reference to the "MECHANISTIC NATURAL SCIENCE" passage on page one hundred eight, there is at the present time no mechanical procedure for coping with the imputations of the three adjectives discussed in the commentary. However, if a technique we have been considering for automatically generating thesauri from dictionaries were employed, and if we had access, through a resource-sharing computer communications networking of very large data bases,[44] to a number of text specific thesauri which were time tagged (that is, the texts were dated), then we might be able to establish automatically the arbitrariness--in fact the incorrectness--of the assignment of the adjectives "mechanistic", "deterministic", and "physiological." But as with many other instances of coping with the kinds of features of a text which we are examining here, the suggestions are only that; the research it would take to put them into effect and then determine whether they worked against the criteria established would be a major job and one whose outcomes are by no means assured.

The passage from page one hundred nine would require the automatic establishment of perspective shift for an elucidation of the passage to be as complete as a human reader might provide. Some of the procedures discussed in Section IV could give us some assistance in locating perspective shift. As to the content or implication of the perspectual alteration, and specifically the adoption of a sarcastic

tone, at a minimum that would include successful analysis of long language strings of the sort discourse analysis is directed toward. Ironies and sarcasm accompanying perspectual transitions imply discontinuities in the logical structure of an argument; more particularly they imply inconsistencies. The approaches of Bohnert and Backer mentioned earlier[45] and others discussed under the headings of semantic and content analytic procedures suggest ways of detecting such discontinua. Again, the effective design and utilization of such a detection process would imply a good deal of original research.

Another approach which has been taken with reference to shifts in tone relates to the consideration of stylistic parameters, where they are taken as not only significant in and of themselves but also as surrogates for other changes in the content of a text.[46] For instance,we have some reason for believing that there may be stylistic shifts which accompany shifts in an individual's perception of the degree of veridicality of assertions which he is making.

Again, with reference to the passage on page one hundred ten, we do not have a computer-based procedure for picking up on the consequences of the particular choice of verbs--in this case "appeared," "was imagined," "was predicted"--particularly when taken in conjunction with the highly contrastive conclusion involved in their negation. It is possible that the rhetorical device involved in the choice of those verbs might be rendered tractable by means of the generation of populations of text-specific thesauri such that, at least on a probabilistic basis, one could assert that a particular selection of

adjectives was exceptionally distant semantically within texts of the type for which the thesauri had been developed. It is also possible that the extensive generation of data which enabled us to achieve certain patterns of normalization would reveal, on more direct statistical treatment, that certain verbs were sufficiently unlikely as co-occurrences with certain populations of nouns and adjectives that there was a special significance to those verbs; that special significance in turn might be derivable from either an empirical regularity or, much more satisfactorily, from a controlling model which itself was culturally specific as to lexical behavior.

As to the passage from page one hundred eleven which is marked by false reification, or hypostasis, it is once more the case that the possible significance or, better, possible reader impact of the use of hypostasis could be caught with programs of the sort proposed in the early work of Ross Quillian[47], insofar as those programs were used to seek out the physical referential implications of natural language statements. That is, a set of programs which consistently requested information as to various physical properties of ostensibily referential discourse and also as to physical actions of the instrumental discourse implicit in all referential discourse (the instrumental discourse would have to be dealt with by some future version of current procedurally-oriented programs), could automatically signal certain deficits as implied by the false lumping together of the referents involved in the hypostasis here. In this procedure a referential lacuna would be detected owing to the absence

of an entity corresponding to a collective noun, while this and other referential discourse would be decomposed into the operations implied by the processes of referencing behavior. Such automatic flagging of a referential lacuna in turn might be related to other parts of a natural language processing model which would characterize such a lacuna as likely to be related to a particularly denominated rhetorical ploy.

Personally, we have also been interested in the problem of the induction of frames of reference from the mechanical examination of specific bodies of text. It is possible that through the comparison of the properties of a semantic space, where the measures of semantic space were established through examinations of characteristic distancings within the standard, 'base-line' thesaurus, one could formulate characterizations of frames of reference and points of view such as are emerging from an examination of this particular passage. There is an intriguing research possibility here but spelling it out in detail would take more space than the present paper allows for.

With reference to the last of the passages from page one hundred eleven, it is possible that contrary to fact assertions or assertions that are probably contrary to fact might with some measure of assurance be detected if we had sufficient access to paraphrasable sense content, which could be validated or invalidated, on a network basis, from large bodies of text.

In conjunction with the idea I have developed elsewhere for a Bicentennial Electronic Encyclopedia (BEE), we have preliminarily considered among implications of that possibility the prospect of detecting statements which are unwarranted by the information at our disposal. Such detection is now principally a matter for traditional scholarship and science in noticing incompatabilities with validated statements appearing in the print medium. But more and more we find the utilization of electronic information storage could merge into predominance, instead of, or along with, the print medium. Provided the structure of the storing is satisfactory--and that is a very significant proviso which carries us into current data structure research problems as well--we could, if those data bases were accessed on a network basis, detect discrepant assertions and what might be unwarranted assertions. Such a procedure might indeed enable us to cope, in some respects at least, with the content of that last passage from page one hundred eleven.

One aspect of that passage, as well as of the passage from page one hundred twelve, is the presence of what might be described as invalid antitheses. Once more the point about invalid antitheses is that they depend on contrary to fact assertions or they imply contrary to fact assertions. One of the ways in which the process of validating assertions may be made more rigorous as well as more automatic, should we choose to do so, is by means of the coordination of content in data bases. That implies, of course, the monumental task--in some loose sense analogous to that of putting together a

major reference work such as the Encyclopedia Britannica or The Oxford English Dictionary--of validating the content in the network files. In some measure one can see, of course, how even that process of validation could be done in a bootstrapping way, such that it was partially automatic. What we are talking of here might be regarded as, in a sense, the present technology's analogue to the Encyclopédie of the 18th century.

In attempting to capture the meaning of "so-called" in the first passage from page one hundred thirteen, if we had in machine readable form The Oxford English Dictionary or its analogue, we might be able to detect a semantic anachronism. Such processes remind one of the work of the great classicist Bishop Bentley. With the cautionary example in mind of Bishop Bentley's 'improving' on Milton's Paradise Lost, one would note that such automatic procedures as we are discussing here are, as has been tacit in our discussion of them, applicable only to text which is being examined with reference to its veridicality. Problems in analyzing the fantastic and the imaginative with reference to criteria other than veridicality pose a fresh set of research tasks. If one had machine access to The Oxford English Dictionary, one could detect many assertions which in part depended for their impact on historical anomalies.

As to the next passage from page one hundred thirteen, the association of "gaily" and "blithely" with the other terms in the passage might be detected as uncharacteristic with reference to their context's having a central tendency of concern manifested by a very different lexicon. That is, establishment of semantic distance by

means of contrast between a text-specific thesaurus and a standard baseline reference thesaurus, such as Roget's International Thesaurus, might enable us at least to catch the unusual quality of what is transpiring in these sentences. What makes the quality unusual as distinct from its being unusual would require again, one might suppose, attention to contrasts in the particular subspaces or subnets within the thesaurus from which vocabularies were being drawn contrastively.

Now we turn to the Helmholtz passage on page one hundred fourteen; we are not commenting on every passage, partly because some of them repeat characteristics of passages discussed earlier. The pejorative conjunction of "Darwinism" and "spell" might be detected on the basis of a thesaural structure showing "Darwinism" as falling within the semantic net of science, positivism, etc., whereas spell might be linked with the mystical and fanciful. In addition, it might be desirable to have available analogues, albeit differently structured, of the W. Nelson Francis post World-War II "Million Word Corpus" of American English[48] for different times and cultural regions. Such corpora might be utilized for the purpose of establishing normalized patterns of language usage, such that noncharacteristic co-occurrences stood out sharply.

The "vogue" of Darwinism and the association of marvels with "Darwinian social science" are two other phrasings from that same page which also might be caught in a net of the sort we have been discussing apropos the Helmholtz passage. The last passage from page one hundred fourteen poses a special problem in that it relates

to the epistemological structure of work done by an author, in this case, of course, Hayes. He uses language as though it had some timeless epistemological significance apart from relationship to particular models as employed by people in seeking to cope with their environments. Thus tacit in the discussion of living organisms is some notion, hinted at through the use of the word, "disclosed," that the terms "motion" and "energy" are meant to stand in some one-to-one relationship to reality and hence are either correctly used or incorrectly used, either true or false. One might rather want to say that Hayes has 'misunderstood' the significance for science of properties of its use of language and seems not to understand that "mechanical," "atomic" or "chemical" do not stand for reality but rather stand for components of models which are used in trying to manage reality. That is to say, their significance is intersymbolic; they serve, for example, as rubrics for sets of human processes represented in symbolism.

Hayes seems to have a kind of windowpane theory of knowledge, and it would not be easy to detect that; or, more generally, while it might not be so difficult to detect that specific orientation the problem is one of getting at the (conceptual) structure of every body of symbolic work--which is what's at issue here in a subtle way. It is a matter about which earlier on we gave some preliminary indication as to a research tack. The indication should make it clear that we are at present in no position to detect automatically the metaphysical assumptions, so to speak, in the work of an individual or set

of workers, much less automatically generate the full structural relationships within their sociosymbolic constructions of realities.

Before closing this section of our essay we will try to provide some further summative comment and also try to look ahead to what may be on the research horizon with reference to solving the kinds of problems we have been seeking to point out here.

It is evident that currently available and prospectively imminent machine-based systems are not capable of coping with most of that response in the reading of a text which we have examined in this test-case from C.J.H. Hayes. That is to say, while we may, in varying measures, have at hand capabilities relating to syntactic parsing, semantic parsing, and, at the extra-sentential level, content analytic structure, as well as some structure-of-discourse capability, we are not now able to deal with the emotion-laden and connotative aspects of discourse.

At the same time we have tried to indicate a variety of tactics which could be employed in trying to cope automatically with the absence of designated agency, the presence of irony, of sarcasm, of various tonal and overtonal properties of the reading of a text, as well as the detection of hypostasis. There are indications, also, that we could, in principle, make progress in locating contrary to fact assertions and major shifts in perspective within natural language strings of considerable length. But while we have indicated tactics which could be utilized toward coping at varying levels of adequacy with these aspects of the reading experience, it is also clear that

we have here indicated in a loose way new research routes to follow without having indicated the expense that it would take to use them, much less the cost of producing fully-implemented, well-defined, self-contained operational systems which could do production runs against large bodies of history of science text. For these phases of the historian's work, then, much would remain to be done if we were to have available a formalized system, even though the outlook for achieving such a system is far from hopeless. It should be pointed out, though, that as with various efforts to deal with restricted universes of discourse, it is the case that we are not sure how high the scientific quality would be of solutions proposed. Many solutions which are today employed in giving us systems for coping with natural language under conditions of tight semantic constraint, are engineering solutions quite lacking in scientific generality. We would say with assurance that some of the solutions we propose would partake of good science, but it is by no means clear that a viable tactic for coping with all of the problems of connotation and the like which we are here concerned with would have significant scientific quality. That would certainly seem to be a possibility with reference to the checking against large linked data bases for evidence of contrary to fact assertion, although it is true that to get the scale of data bank information that is required does imply a modularization of data bases and a dramatic improvement in their general structural properties beyond what we currently have.

In Section V of this essay there is discussion now and again of some of the more promising research strategies and tactics for

improving the armamentarium of devices now available to us for understanding natural language. And, of course, that is also true in this section as well. If we examine current natural language research in progress we may feel that language-understanding systems are in better shape as an active research domain than is true with, for example, information systems, where there seems little evidence of the progress or ferment which one would hope for in an important research specialty.

Beyond the instances of significant work in progress already cited perhaps one might also note that certain current research on man machine dialoguing, as for instance with computer-based consultant information systems, holds some promise for also being of assistance to us.[49] So, too, with the production rule approach employed by many of the people working on artificial intelligence in the medical, biomedical, and chemical sciences.[50] Through the utilization of formats heavily involving statements of the If A, then B type, much good work is being done as to the process of logical inference, as for example in conjunction with medical diagnosis and also with the formulation of answers to the riddles of the structures of large molecules. These approaches are significant for helping us through inference chaining to get at the meaning of a passage.

There is some feeling that formal research relevant to context free systems has especially great promise and that augmented context free systems may be one of the likely ways of automatically coping with some of the complexities in natural language discourse[51], just as there are those who are currently very interested in the

implications of Montague grammars for that purpose.

Still, we do not feel that there is any prospect of the formulation of a powerful, new paradigm which would give us far greater control than we have now over the complexities of natural language. It is our general conviction that such a new paradigm could emerge from efforts to relate basic central nervous system research to language research. Heretofore there has been much speculation about how the mind may be supposed to work as a function of the language strings it produces, but it should be emphasized that it is very largely speculation. To actually relate language understanding back to the neurological correlates of language generation has, in our judgment, extremely exciting potential. But it has yet to be done. Still, there are evidences, such as, for instance, the recent article in Science by Francis Schmitt[52] and others of the growing implications of achievements on the neurological sciences front for the understanding of overt behavior, such as language-responding behavior and language-generating behavior.

One of the interesting questions that emerges from attempts to build machines to respond intelligently to language is the question of what happens when connotation isn't perceived. A great deal of cultural learning consists in coming to assign attitudes in a conventionally expected way to experience, and especially to its symbolization. Not only is that one of the dominant features of cultural life but it is one of the principal ways in which class-related stigmata are produced. As a function of social position

people learn to assign connotation differentially. It has long seemed to us that it is at least an open question as to whether the replication of the connotative dimension in machine-based systems is without its disadvantages. In fact, in some of our earliest work we argued that a reason for close attention to attempting to simulate the functions of a literary critic by machine was to make far more explicit the, so to speak, rules of the game used by a literary critic. Then a student would not find himself in a situation of ostensibly learning overt symbolic manipulative procedures, when in fact he was being rewarded for acquiring a tacit skill at generating connotation in ways acceptable to a teacher.

Apart from any unwholesomeness that such arrangements impart to pedagogic processes--and there are many who feel that is a central issue of education in our time, which partially surfaced in the 60's--it also may be the case that the acquisition of certain sorts of connotation patterns are, for a society as a whole, rather dysfunctional. They are disadvantageous insofar as they unduly limit the range of response, and hence of understanding, and hence of possible action, at times when the established ways of relating to reality may be inadequate, even against the higher priority preferential behaviors of members of a given culture. So it may be argued that learning connotation can be a restriction on a more noetic response to verbal and other symbolic stimuli. And it may be that learning connotation contributes to the deterioration which each human central nervous system experiences in becoming more and more

rigid, less and less capable of useful variety of response, as a function not only of such time-frames as may produce the expectation of arterial sclerosis, but even within such time-frames as produce most of the formal and informal enculturation and socialization that take place up until, let us say, early middle age. We need seriously to consider, then, whether trying to make computers emulate people with respect to the human use of connotation, as well as in other respects, is the advantageous course.

As we made clear earlier, fundamentally we feel that complementarity to human functioning rather than its emulation is the preferable course.

In pursuit of further understanding of how computer-based language-analytic systems may be developed and employed to complement the work of the historiographer, we hope to carry further work we had begun earlier on Hume's History of England. Walter Sedelow has done an examination of Hume's History with reference to its lexical, rhetorical, and mythic components. In a separate study we have run the VIA programs against large blocks of the History's text. What we now hope to do is to bring out a study showing, in parallel, what was revealed by a careful, structured human reading of the text and what a machine-based reading produced. There is no doubt at all but what certain phenomena in a human reading are not captured by current machine procedures, but what is also apparent is that there are associative networks within the text which bind it together and create impressions of unity and style that are more fully captured even by rather elementary programs than they are by a careful reader.

VII.

There is a vulgar error about science--which is that science is essentially a matter of measurement and, hence, of quantification. Similarly, there is a vulgar error about mathematics--which is that mathematics is essentially sophisticated arithmetic. Now, like many vulgar errors, there is a great deal more than just a grain of truth in those notions. Nonetheless, these views are erroneous.

As we consider the utilization of computers and computer science in conjunction with the study of the history of science, we may feel that we are examining another chapter in the long human story of imitation and frustrated innovation. It is a frequent phenomenon that the work of those who would imitate is marked by some deficiency, some simplification of complexity which has omitted critically important properties. Often that paleness or weakness in an imitative procedure is associated with some lack of grasp on the process as a whole which is being emulated and, more particularly, there may be a lack of relating the derivative structure as a whole to the accomplishment of a goal.

Thus, there is a danger in much effort by social scientists and historians to model themselves upon the so-called 'hard sciences' in that they may relate to the surface of science rather than to its present depth. There is in this statement no implicit criticism of the effort to produce a more quantified history of science. All other things being equal, that is a very attractive goal insofar as it gives us more precisely formulated assertions in place of weaker assertions.

And with these more precisely formulated assertions we may be able to detect relationships among dimensions and processes in the science generative process that would not have been detected otherwise and that through enhancing our understanding of science as a social and cultural activity--as the activity of groups of specialized language users[53]--would enable us to so better understand science itself that we could create the conditions for generating new scientific knowledge more readily when we need it.

Any compulsive imitationof an existing product or process while utilizing a new technology can lead to a failure to utilize well that new technology. The history of technology is replete with instances of new technologies not being effectively employed in their early phases because their own distinctive properties were not exploited relative to their cultural and technological environments. Rather, the new technologies were measured as to success in imitation of some other, earlier technology. There is a danger that historians of science for their own métier will not mobilize the most fundamental technological and even scientific knowledge in ways which are distinctively appropriate to their own requirements. The danger is not only that of educators or social scientists seeking to become in some cases more scientific by compulsive counting. Even more fundamentally, there is the danger that in the application of the computer we will try to imitate the behavior of people, as distinct from imitating the behavior of other types of scientists. That is to say, just as it may be unfortunate for one group of scholars or scientists

to rather heedlessly seek to convert themselves into versions of scientists who are prestige-laden, so, too, it may be a major error of a comparable sort to attempt to employ computers to imitate people.

Earlier on we pointed out that while one might usefully think of various specialized forms of artificial intelligence (AI), such as artificial medical intelligence or even, in this case, artificial science-historical intelligence, the general notion of artificial intelligence was adversely affected by a failure to realize that what was being simulated was not human CNS functions, or even specialized subsets of them appropriate to particular types of scientific work, but rather symbolic outputs by human beings whilst performing certain roles. There are certain scholarly or scientific activities which the historian of science may find there are algorithms for handling on a computer more satisfactorily than a human being can perform them. In such cases, economics allowing, we may wish to use the computer to accomplish them. We need, though, to be wary of the possibility that what we are doing is using the computer not in so imaginative, original, and effective a way as we might, but instead rather phobically copying human behavior by machine--just as much human behavior itself is phobically imitative rather than responsive to its environment, and to such a degree that a great deal of human culture (technology included), amounts to a gigantically sanctioned collection of phobias dignified by the name of culture.

As we look to horizons of the future we may well find, for example, that it was fundamentally wrong-headed to design computer-

based systems to cope with connotation and the like on the order of the way in which an historian of science might in reading a text. That is to say, we might ask whether we want to imitate connotation, whether there are fundamental advantages to not coping with it, whether there may be certain fundamental advantages to utilizing language in a connotation-free mode. We may suspect that connotation is used only when precision of a comparatively effective sort in the denotative mode is not available. One might hypothesize that, ceteris paribus, there is an irreversible progression to effectiveness, such that the imprecision, the unclarity, the lack of operationally precise consequence of connotation is in the nature of the case abandoned when there are denotatively effective alternatives. It would seem as though it might be argued that the history of science itself tends to bear out such a contention, in that for any given individual (except during a condition of such sort that we might label it psychic deterioration) as well as for sets of individuals, knowledge replaces superstition, ignorance and the like, and not the reverse. Knowledge grows and science cumulates. In understanding how that comes to pass we note an important consideration: that the conditions generated through the utilization of knowledge are more stable than those that come to pass where the human involvement is less knowing, just because in its nature knowledge means a more effective understanding of self or environment, or self-in-environment, which is to say a more effective knowledge of nature. The essentially system stabilizing quality of the application of intelligence through knowledge was

notably clarified through the celebrated book Design for a Brain, by the late W. Ross Ashby. In his later Introduction to Cybernetics he carried those ideas further.[54]

It might, then, be argued that the connotation-free response of a machine-based system for relating to materials in the history of science could give us a way of perceiving science in which the essentially scientific stood forth more clearly. For example, once content analytic and denotative semantic systems appropriately located the scope of response to symbolic materials as to domain of action, and instrumentally oriented systems converted the operations components of text into their implied actions, we would then have the strictly scientific component of the larger process that we more loosely call science well isolated. That might be a very considerable advantage. We have found an interesting resonance to this point of view in the recent work, Introduction to Metascience, An Information Science Approach to Methodology of Scientific Research by Thomas J. Hickey.[56] His case study, a formal explication of Keynesian theory, is an entrancing piece of work. His book, which grew out of work in the graduate program in the philosophy of science at Notre Dame, reveals an interesting way of formalizing the history of science. While doing so he implicitly shows how the history of science in the course of such formalization may contribute to the enhanced effectiveness of science itself, by means of computer-implicated procedures. After all, in a world in which more and more traditional academic processes are being questioned, it may be useful to be able

to argue more forcefully than formerly for the tacitly assumed relevance of the history of science for science itself.

It is desirable to note here that a question is being raised rather than an answer given. The question is not whether it is desirable to be able to understand the process of connotation as it enters into the human reading experience. We take it for granted that is a worthy and appropriate goal for scientific investigation. We do ask, though, whether for understanding the structure among the components in the up-building of positive scientific knowledge there might be advantages in being able at times to examine texts with reference only to their instrumental significance, and utterly without regard to their connotation. It is unthinkable that a human could ever do that on his own--although there would no doubt be degrees of success in trying. But it is not unthinkable that a machine might be able to do so, and that on occasion that would be usefully complementary to human readings.

There is the distinct possibility that the history of science itself may be fundamentally characterizable by a succession of state-shifts that would provide the macrostructure within which paradigm shifts attended to by Thomas Kuhn might be found to fit: state-shifts as the paradigm for paradigm shifts. If it is the case that we find that a great deal of detail in the history of science, and in science itself, amounts, in an information theoretic perspective, to noise in the system, or in a mechanical perspective to friction in the machine, we may discover that it is in the formalization of the

transformation processes that carry us from one paradigm to another that the essentials of both science and the history of science are located. Should we find that is the case, then it would be interesting to note that science itself--not surprisingly within the context of expectations of parsimony and of continuities across the processes of nature--was in a fundamental sense a natural process, too. For it is certainly the case that we are ever more able to put together pieces of our understanding of nature within the structure of notation and concept that are employed in the study of abstract machines.

As you know, the computer is, formally, a machine, a finite state sequential machine, just as the software which runs on it is a machine. We find more and more that computer science is central to the sciences, and not just because of the use of particular physical computers in handling numerical data reduction. Computer science is central for the more essential reason that the conceptual structure of computer science--that is to say finite state sequential machine theory--can be a master paradigm for scientific knowledge at large. In that connection, at the more philosophical level the recent book Chance and Necessity by the late French chemist Jacques Monod [56] provides a general frame-of-reference orientation, while the recent study of Arthur W. Burks, Chance, Cause, Reason[57] carries us a step closer to the operational realization of the concepts involved and books by Burks' associates Holland and Ziegler carry us even further.[58]

We have long known that there is nothing inevitable about the individual scientific disciplines and specialties as we know them,

and their mutual relations. They metamorphose through time in any given culture, and from one culture across to another we cannot strictly map them. Illustratively, in classical China there was no clean analogue to what we call history. It may come to pass that the breakout into academic specialties within science as we know it today will be fundamentally altered through the emergence of a powerful generalized science to be applied to specific but changing sectors of reality by generic scientists--rather than by people who are so specialized as are the current practioners of academic subspecialties in the sciences. A speedup in the growth of scientific knowledge alone would be enough to disrupt radically the existing structure of disciplines and subdisciplines and perhaps even to push us to a generic role for the practicing scientist. If that were to happen, surely it would be the case that an increased component of the scientist's role would be the formal analysis and generation of symbolic structures suited to the kinds of realities he was seeking to cope with. In any event, computers and computer science may also contribute in the future to a change in the role of the historian of science, not only because through theoretical computer science science itself may well change, but also because as in fundamental ways the history of science itself becomes more scientific, the boundary conditions between the history of science and science itself may tend to fade.

FOOTNOTES

1. For related comments on the nature of historiography, see Walter A. Sedelow, Jr., "History as Language," Computer Studies in the Humanities and Verbal Behavior, Vol. I, No. 4, December, 1968, pp. 183-190.
2. See, for example, E. A. Feigenbaum, B. G. Buchanan, and J. Lederberg, "On Generality and Problem Solving: A Case Study Using the DENDRAL Program," Machine Intelligence 6, ed. B. Meltzer and D. Michie, Edinburgh University Press, 1971.
3. For a more extended overview of the research and research goals of this group, see Proceedings, 1975 AIM/SUMEX Conference, Rutgers University, New Brunswick, New Jersey.
4. For a discussion of some aspects of formalism in the humanistic disciplines, see John B. Smith, "Computer Criticism," Computers in Language Research: Formalization in Literary and Discourse Analysis, ed. Sally Yeates Sedelow and Walter A. Sedelow, Jr., Mouton, in press. See also Walter A. Sedelow, Jr., "Bibliography for a Science of Language," Computer Studies in the Humanities and Verbal Behavior, Vol. V, Nos. 1-4, 1974, pp. 1-91.
5. Walter A. Sedelow, Jr., "Science and the Language of History," Behavioral Science, Vol. 2, No. 1, 1957.
6. C.J.H. Hayes, A Generation of Materialism in "The Rise of Modern Europe" Series, ed. William L. Langer, New York, 1941.
7. See W. A. Sedelow, "History as Language," op. cit.

8. Ralph Grishman, "A Survey of Syntactic Analysis Procedures for Natural Language," Courant Computer Science Report Number 8, August, 1975, Courant Institute of Mathematical Sciences, New York University.

9. David B. Benson, "Formal Languages vis-a-vis 'Natural' Languages," Computers in Language Research: Formal Methods, ed. Sally Yeates Sedelow and Walter A. Sedelow, Jr., Mouton, in press.

10. For additional comments on these examples as well as others, see Sally Yeates Sedelow, "Analysis of 'Natural' Language Discourse," AFIPS Conference Proceedings, Vol. 45, 1976, AFIPS Press, Montvale, New Jersey, pp. 1027-1034.

11. W. A. Woods, R. M. Kaplan, B. Nash-Webber, The Lunar Sciences Natural Language Information Systems: Final Report, BBN Report No. 2378, June 15, 1972, Bolt, Beranek, and Newman, Inc., Cambridge, Mass. See also, W. A. Woods, "Transition Network Grammars for Natural Language Analysis," Communications of the Association of Computing Machinery, Vol. 13, No. 10 (Oct. 1970), pp. 591-606.

12. David Benson, op. cit.

13. Terry Winograd, Understanding Natural Language, Academic Press, New York, 1972.

14. David Benson, op. cit.

15. Benson, ibid.

16. Benson, ibid.

17. Benson, ibid.

18. F. B. Thompson, et al., "REL: A Rapidly Extensible Language System," Proceedings 24th National Conference Association of Computing Machinery, pp. 399-417; F. B. Thompson and B. H. Dostert, "The Future of Specialized Languages," AFIPS Conference Proceedings, Vol. 40, 1972, pp. 313-319.

19. Benson, op. cit.

20. Sedelow and Sedelow, ed., Computers in Language Research: Formal Methods.

21. Montague Grammar, Academic Press, New York, 1976.

22. See Roger C. Schank, "Conceptual Dependency: A Theory of Natural Language Understanding," Cognitive Psychology, Vol. 3, 1972, pp. 552-631. See also, David E. Rumelhart, Peter H. Lindsey, Donald A. Norman, A Process Model for Long-term Memory, May, 1971, Center for Human Information Processing, University of California, San Diego.

23. For an overview of earlier computer-based work in content analysis, see George Gerbner, et al., ed., The Analysis of Communication Content: Developments in Scientific Theories and Computer Techniques, John Wiley & Sons, Inc., 1969.

24. A discussion of recent quantiative work in what might be termed historiography is provided by Régine Robin in "On Quantitative Research Outside the History of Science: The Case of Discourse Analysis," Working Papers (Roger Hahn, editor) for The International Symposium on Quantitative Methods in the History of Science, August 25-27, 1976, Berkeley, California.

25. Third Edition, Thomas Y. Crowell Company, New York, 1962.

26. For detailed discussion of aspects of this work see Sally Yeates Sedelow and Walter A. Sedelow, Jr., "Categories and Procedures for Content Analysis in the Humanities," in George Gerbner, op. cit., pp. 487-499; Sally Yeates Sedelow, "Analysis of 'Natural' Language Discourse," op. cit.; and research reports by Sally Yeates Sedelow, et al., for the years 1969-1970 (DDC #AD711-643), 1971 (DDC #755-134), 1971-1972 (DDC #AD752-508), 1972-1973 (DDC #AD774-692), and 1973-1974 (DDC #ADA002-463). The reports are available through the National Technical Information Service (U.S. Dept. of Commerce), Springfield, Virginia.

27. See Sally Yeates Sedelow, "Analysis of 'Natural' Language Discourse," op. cit.

28. David Benson, op. cit.

29. Benson, ibid.

30. Benson, ibid.

31. See Sally Yeates Sedelow and Walter A. Sedelow, Jr., "A Preface to Computational Stylistics," The Computer and Literary Style, ed. Jacob Leed, Kent State University Press, 1966, pp. 1-13; see also by the same authors, "Stylistic Analysis," in Harold Borko, ed., Automated Language Processing: The State of the Art, John Wiley & Sons, 1967, pp. 181-213.

32. Jane Robinson, "A Tuneable Performance Grammar," Stanford Research Institute, Artificial Intelligence Center, Technical Note 112, September, 1975, p. 1; additional remarks from Sally Yeates

Sedelow, "Analysis of 'Natural' Language Discourse," op.cit., p. 1028.

33. Technical Report Number 1, NSF Grant GS-3180, Cornell University, 1972.

34. Grimes, op. cit., pp. 116-117.

35. Grimes, op. cit., pp. 281-294.

36. See Sedelow and Sedelow, "A Preface to Computational Stylistics," op. cit., and "Stylistic Analysis," op. cit.

37. John B. Smith, "Image and Imagery in Joyce's Portrait, eds. Stanley Weintraub and Philip Young, Directions in Literary Criticism, University Park, Pa., 1972, pp. 220-27; see also by same author, "Computer Generated Analogues of Mental Structure from Language Data," Proceedings of IFIP '74, The Hague, 1974, pp. 842-5.

38. op. cit.

39. Robert P. Abelson and J. Douglas Carroll, "Computer Simulation of Individual Belief Systems," American Behavioral Scientist, Vol. VIII, No. 9 (May, 1965), pp. 24-30; Robert P. Abelson, "Modes of Resolution of Belief Dilemmas," Journal of Conflict Resolution, No. 3, 1959, pp. 343-352.

40. Winograd, op. cit.

41. "An Overview of Networking-for-Science," National Science Foundation, Washington, D.C., mimeographed, 1975.

42. M. R. Quillian, "Semantic Memory," Marvin Minsky, ed., Semantic Information Processing, M.I.T. Press, Cambridge, Mass., 1968.

43. Philip J. Stone, et al., The General Inquirer, M.I.T. Press, Cambridge, Mass., 1966.

44. Walter Sedelow, "Network Research: Some Issues," NSF Director's Program Review, Computer Research, National Science Foundation, Washington, D.C., 1975, pp. 15-22. See also, W. Sedelow, "Networks for a Nation's Scientific Disciplines," 3rd International [Toronto] Conference on Computer Communications, International Council on Computer Communication, Washington, D.C., 1976; and W. Sedelow, "Some Implications of Computer Networks for Psychology, Behavior Research Methods and Instrumentation (Psychonomic Society), Vol. VIII, #2 (May 1976), pp. 218-222. As to very large data bases, see Proceedings of the International Conference on Very Large Data Bases, Douglas Kerr, ed., Association for Computing Machinery, New York, 1975.

45. Bohnert and Backer, op. cit.

46. Sedelow and Sedelow, "A Preface to Computational Stylistics," op. cit., and "Stylistic Analysis, op. cit.

47. M. R. Quillian, op. cit. See also, M. R. Quillian, "The Teachable Language Comprehender," Communications of the Association for Computing Machinery, Vol. 12, 1969, pp. 459-476.

48. For a description of this corpus, see Proceedings of the IBM Literary Data Processing Conference, Yorktown Heights, New York, 1964.

49. Sites where such research is in progress include the Information Sciences Institute, University of Southern California, and the Stanford Research Institute, Menlo Park, California.

50. 1975 AIM/SUMEX Conference, op. cit.

51. See Benson, op. cit.

52. Francis O. Schmitt, et al., "Electrotonic Processing of Information by Brain Cells: Recent Research Augurs an Important Role for Neural Local Circuits in Higher Brain Functions," Science, Vol. 193, 9 July, 1976, pp. 114-120. See also, M. Conrad, W. Guttinger, and M. Dal Con, edd., Physics and Mathematics of the Nervous System, Proceedings of a Summer School, Trieste Italy, Aug. 1973, Springer-Verlag, New York, 1974, and its review by David Marr in Science, Vol. 190, 28 November, 1975, pp. 875-876.

53. See W. A. Sedelow and S. Y. Sedelow, "Language and People, Modelling and Science," Computers in Language Research: Formal Methods, W. A. Sedelow and S. Y. Sedelow, ed., Houton & Co., The Hague, in Press.

54. W. Ross Ashby, An Introduction to Cybernetics, New York, 1957; and Design for a Brain, London, 1954.

55. Oak Park, Illinois, 1976.

56. Jacques Monod, Chance and Necessity, New York, 1972.

57. University of Chicago Press, Chicago, Illinois, 1975.

58. John Holland, Adaptation in Natural and Artificial Systems, University of Michigan, Ann Arbor, Michigan, 1975; and Bernard P. Ziegler, Modelling and Simulation, Wiley, New York, 1976.