DOCUMENT RESUME

ED 028 633 24 EM 000 317

By-Page, Ellis B.; Paulus, Dieter H.
The Analysis of Essays by Computer. Final Report.
Connecticut Univ., Storrs.
Spons Agency-Office of Education (DHEW), Washington, D.C. Bureau of Research.
Bureau No-BR-6-1318
Pub Date Apr 68
Contract-OEC-16-001318-1214
Note-280p.
EDRS Price MF-$1.25 HC-$14.10
Descriptors-*Computer Assisted Instruction, Content Analysis, Educational Diagnosis, Educational Testing, *Essays, Essay Tests, Information Storage, Instructional Technology, Predictive Ability (Testing), Predictive Measurement, Psychometrics, *Simulation, Statistical Analysis, *Structural Analysis, Student Writing Models, Time Sharing, *Writing Skills
Identifiers-PEG, *Project Essay Grade

This study aimed at expanding a new field of educational measurement, by investigating the feasibility of using computer programs for the automatic analysis and evaluation of student writing. Essays written by secondary students in their English classes were rated by multiple independent judges on a number of traits usually considered important: content, organization, style, mechanics, creativity, and overall quality. The essays were key-punched for input to the computer. Computer programs were written to analyze the essays, performing many tests and list lookup procedures, and producing a profile of "proxes" (variables believed to be approximations of important dimensions of the essays). These proxes were then combined through multiple regression to optimize the prediction of the expert judgments. Across various essays, judges, students, and traits, the computer performed about as accurately (in predicting the expert group) as did the typical human judge. Many other dimensions of the problem were examined, including the use of cliches, passive verbs, and syntactic parsing. A plan of attack was outlined for future investigators. (Author)

Final Report

Project No. 6-1318
Contract No. OEC-16-001318-1214

Principal Investigator:
Ellis B. Page

THE ANALYSIS OF ESSAYS BY COMPUTER

by

Ellis B. Page

and

Dieter H. Paulus

The University of Connecticut
Storrs

April, 1968

U. S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

Final Report

Project No. 6-1318
Contract No. OEC-16-001318-1214

Principal Investigator:
Ellis B. Page

THE ANALYSIS OF ESSAYS BY COMPUTER

by

Ellis B. Page

and

Dieter H. Paulus

The University of Connecticut
Storrs

April, 1968

The research reported herein was performed pursuant to a contract with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

TABLE OF CONTENTS

	Page
TITLE PAGE	i
TABLE OF CONTENTS	ii
PREFACE BY THE PRINCIPAL INVESTIGATOR	v
THE TEXT	
CHAPTER I: INTRODUCTION	1
The Practical Background	1
The Input Question	3
The Theoretical Background	4
Related Research	6
Background Disciplines	6
Psychometrics	6
Linguistics	7
Curriculum	7
Automatic Language-Data Processing	8
Statistical Methodology	8
Computer Technology	9
Objectives of the Research	10
CHAPTER II: THE BASIC DESIGN	13
Rationale	13
CHAPTER III: THE INITIAL PROXES	19
Sampling	19
Hypotheses and Proxes	21
The Computer Program	31
Preparation of the Text	38
Summary	39
CHAPTER IV: PREDICTING OVERALL QUALITY	40
Human Ratings	40
Overall Prediction of the Proxes	46
Correlation With the Criterion	46
Multiple Regression	51
The Paulus Tables	53
Use of the Tables	59
Reliability of Proxes	60
Human and Machine Judgments	62
Human vs. Machine "Validity"	63
Human vs. Machine Accuracy	65
Using One Essay's Proxes for Another Essay's Criterion	66

TABLE OF CONTENTS (Continued)

	Page
CHAPTER IV (Continued)	
Cross-validation with Same Essays	69
Practical Implications	71
Summary	72
CHAPTER V: PREDICTING A PROFILE OF RATINGS	73
The Sample	73
The Rating Session	75
The Rating Criteria	76
Contribution of the Proxes	83
The Uniqueness of the Traits	92
Judge Viewpoints	96
Trait Prediction by Machines	100
Summary	105
CHAPTER VI: PROBLEMS OF STATISTICAL IMPROVEMENT IN PREDICTION	107
The Problem of Linearity	107
Interactions	109
Transformations	116
Discussion	122
Summary	125
CHAPTER VII: PHRASE LOOKUP AND ITS IMPLICATIONS	126
The Phrase Look-up Procedure	126
An Application to Cliches	130
Background on Cliches	131
A Search for Psychological Characteristics	134
Correlative Conjunctions	139
Verb Constructions	141
Parenthetical Expressions	142
Summary	149
CHAPTER VIII: ON-LINE ANALYSIS AND FEEDBACK	150
Background	151
The Program	152
Summary	157

TABLE OF CONTENTS (Continued)

	Page
CHAPTER IX: CONCLUSIONS AND IMPLICATIONS	163
Summary of Work Completed	164
Rationale	164
Findings	165
Some Work Related to the Project	169
Journals	169
Societies	169
Textbooks	170
Other Books	171
Recent Related Work	172
Parsing	173
Multiple Path System	173
Discourse Analysis	185
Transformational Grammar	186
Semantics	186
Future Work in Essay Analysis	189
Need for Flexibility	189
Grading of Content	191
Further Analysis of Style	194
Hypothetical Complete Essay Analyzer	194

APPENDICES

APPENDIX A.	Source Program for Essay Analysis	199
APPENDIX B.	Statistical Tables for Multiple Regression	243
	Table IV-11 (A)	243
	Table IV-11 (B)	244
	Table IV-11 (C)	245
	Table IV-11 (D)	246
	Table IV-11 (E)	247
	Table IV-11 (F)	248
	Table IV-11 (G)	249
APPENDIX C.	Source Listing for PHRASE	250
APPENDIX D.	PL/I Source Program for PARSE	252
REFERENCES		264

TABLES

		Page
TABLE IV-1	HUMAN JUDGE AND JUDGE PAIR CORRELATIONS FOR ESSAY C QUALITY WITH PREDICTIONS FROM ESSAY D	42
TABLE IV-2	MEANS AND STANDARD DEVIATIONS OF THE PROX SCORES	44
TABLE IV-3	PROXES USED TO PREDICT A CRITERION OF OVERALL QUALITY (ESSAY D)	45
TABLE IV-4	THE DIRECTION OF CORRELATIONS OF PROXES WITH THE CRITERION: PREDICTED AND OBSERVED FREQUENCIES	47
TABLE IV-5	INTERCORRELATIONS OF THE PROXES FOR ESSAY D	48
TABLE IV-6	INTERCORRELATIONS OF THE PROXES FOR ESSAY C	49
TABLE IV-7	PROXES USED TO PREDICT A CRITERION OF OVERALL QUALITY (ESSAY C)	52
TABLE IV-8	MINIMUM MULTIPLE CORRELATIONS REQUIRED FOR SIGNIFICANCE AT THE .05 LEVEL	54
TABLE IV-9	MINIMUM MULTIPLE CORRELATIONS REQUIRED FOR SIGNIFICANCE AT THE .01 LEVEL	55
TABLE IV-10	MULTIPLE REGRESSION COEFFICIENTS CORRECTED FOR ATTENUATION CAUSED BY CRITERION UNRELIABILITY	57
TABLE IV-12	ESSAY D PROXES USED TO PREDICT AN ESSAY C CRITERION	68
TABLE IV-13	CROSS-VALIDATION COMPARISON OF THE COMPUTER WITH FOUR HUMAN JUDGES	70
TABLE V-1	INTERCORRELATION MATRIX FOR 32 RATERS FOR TOTAL SCORES	81
TABLE V-2	MEANS AND STANDARD DEVIATIONS OF THE FIVE TRAITS AND THEIR AVERAGE	82
TABLE V-3	PROX CONTRIBUTION TO THE PREDICTION OF IDEAS OR CONTENT	84
TABLE V-4	PROX CONTRIBUTION TO THE PREDICTION OF ORGANIZATION	85
TABLE V-5	PROX CONTRIBUTION TO THE PREDICTION OF STYLE	86
TABLE V-6	PROX CONTRIBUTION TO THE PREDICTION OF MECHANICS	87
TABLE V-7	PROX CONTRIBUTION TO THE PREDICTION OF CREATIVITY	88

TABLES (Continued)

		Page
TABLE V-8	PROX CONTRIBUTION TO THE PREDICTION OF AVERAGED RATING ACROSS 5 TRAITS	89
TABLE V-9	INTERCORRELATIONS OF THE TRAITS OF JUDGED ESSAYS	93
TABLE V-10	TRAIT BY ESSAY INTERACTION	95
TABLE V-11	COMPUTER SIMULATION OF HUMAN JUDGMENTS FOR FIVE ESSAY TRAITS	103
TABLE VI-1	RANK-ORDERING OF PREDICTOR VARIABLES	114
TABLE VI-2	MULTIPLE REGRESSION EQUATION FOR LINEAR TERMS AND INTERACTIONS	115
TABLE VII-1	TRITE PHRASES FOUND IN HIGH SCHOOL ESSAYS	133
TABLE VII-2	CORRELATION OF FIVE MAJOR TRINS WITH "OPINIONATION," "VAGUENESS," AND "SPECIFICITY"	136
TABLE VII-3	CORRELATIONS OF VOCABULARY MEASURES WITH "OPINIONATION," "VAGUENESS," AND "SPECIFICITY"	138
TABLE VII-4	DISCOVERED FREQUENCIES OF CORRELATIVE CONJUNCTIONS	140
TABLE VII-5	PARENTHETICAL EXPRESSIONS USED	145
TABLE VII-6	MEAN GRADE DIFFERENCES BETWEEN THE PARENTHETICAL AND NON-PARENTHETICAL GROUPS	146
TABLE VII-7	CORRELATIONS BETWEEN POSITIONS OF PARENTHETICAL EXPRESSIONS AND STYLE	147
TABLE IX-1	THE RELATION OF COMPUTER PARSING TO JUDGED GRAMMATICALNESS OF STUDENT SENTENCES	182
TABLE IX-2	MACHINE PARSING PERFORMANCE OF GRAMMATICAL AND UNGRAMMATICAL STUDENT SENTENCES	184

FIGURES

		Page
FIGURE II-1:	POSSIBLE DIMENSIONS OF ESSAY GRADING	14
FIGURE III-1:	GENERAL FLOW CHART FOR FIRST PROGRAM PROJECT ESSAY GRADE (ESSAY ANALYSIS)	34
FIGURE V-1:	CRITERIA FOR RATING THE ESSAYS	78
FIGURE VI-1:		108
FIGURE VI-2:	VARIABLE NUMBER 8	117
FIGURE VI-3:	VARIABLE NUMBER 15	118
FIGURE VI-4:	VARIABLE NUMBER 22	119
FIGURE VI-5:	VARIABLE NUMBER 23	120
FIGURE VI-6:	VARIABLE NUMBER 29	121
FIGURE VI-7:	SAMPLE OUTPUT FROM TRANSFORMATION PROGRAM	123
FIGURE VIII-1:	SAMPLE COMPUTER OUTPUT	158
FIGURE IX-1:	COMPUTER LISTING OF HOMOGRAPHS FROM THE PARSING DICTIONARY FOR A STUDENT SENTENCE	174
FIGURE IX-2:	FIRST COMPUTER ANALYSIS OF SYNTAX OF A STUDENT SENTENCE	176
FIGURE IX-3:	LATER COMPUTER ANALYSIS OF SYNTAX OF A STUDENT SENTENCE	178
FIGURE IX-4:	STATISTICAL INFORMATION PRODUCED BY THE MULTIPLE PATH SYNTACTIC ANALYZER (SYNTAX DIAGNOSIS)	180
FIGURE IX-5:	FURTHER STATISTICAL INFORMATION PRODUCED BY THE MULTIPLE PATH SYNTACTIC ANALYZER (SYNTAX SUMMARY)	181
FIGURE IX-6:	HYPOTHETICAL COMPLETE ESSAY ANALYZER	195

PREFACE BY THE PRINCIPAL INVESTIGATOR

This document, The Analysis of Essays by Computer, is primarily intended as the Final Report for the United States Office of Education, for a research contract which supported us during 1966 and 1967. Yet it also represents the first summary statement of all of the work undertaken since early 1965 at the University of Connecticut in such essay analysis, and in the simulation of human rating behavior.

It is difficult to trace the genealogy of any idea, let alone one as interdisciplinary as that underlying the present work. The notion of computer analysis of essays began to seem conceivable, following an invitational conference on data banks, led by John B. Carroll at Harvard University in December, 1964. My own experience had included work in many of the contributing fields, so that the manipulation of language, as described by Philip Stone and others there, drew together many threads into an eventually engrossing central problem.

From the moment of conception, this work has owed much gratitude to a succession of able and helpful people. J. A. Davis was immediately encouraging, as were Allan B. Ellis, William Asher, Dexter Dunphy, and Marshall Smith. John Duggan and John Valentine, of the College Entrance Examination Board, helped greatly in arranging almost immediate financial support. All that we did then and later owed much to this prompt generosity of the CEEB, and this report will also serve as the most unified summation of the earliest work done under that support.

Other generous support, supplementary to that of the U.S. Office of Education, has been given by National Science Foundation, through its partial funding of the University of Connecticut Computer Center. Furthermore,

the Massachusetts Institute of Technology was very helpful in supporting me as New England Visiting Scientist to their Computation Center during 1966-67. Finally, the University of Connecticut Research Council has given prompt aid at crucial times.

It would be impossible to list everyone who has been helpful with this Project, and there are sure to be important and unintentional omissions. Here at Connecticut, many ideas were early discussed with Herbert Garber, then with us in the Bureau of Educational Research, with Arthur Daigon, with Charles McLaughlin, and with Kenneth G. Wilson. These have all served as consultants for brief or longer periods of time, and many have contributed ideas or insights which, because of the nature of this report, are not acknowledged explicitly in the text. From the start, the Project had, as principal programmers, Gerald and Mary Ann Fisher. Mr. Fisher has been a consultant and, for the year 1967-68, a Research Associate with us. The programs from this employment have plainly been of central importance to the work.

In mid-1966 Dieter H. Paulus joined the Bureau of Educational Research, and has in many ways contributed richly to the work since that time. His various contributions are mentioned often in the text and he is second author of this report and partner in the on-going work.

Others who helped here in the Bureau of Educational Research were Miss Louise Patros, together with her willing staff of Mrs. Helen Ring, Miss Evelyn Haddad, and Mrs. Katherine Showalter. To Miss Patros much gratitude is owed for office management functions so important to a large research, and to all we are grateful for the preparation of this manuscript. Some of the research detail was carried out by graduate students here in the Bureau. Their names are mentioned in the text, together with their contributions,

wherever these are included in the report. Among these, Donald Marcotte made contributions which were clearly outstanding.

During the work we have consulted many scholars from other institutions, formally or informally, and some of them should surely be listed here: Walter and Sally Y. Sedelow, Robert Stake, Paul Lohnes, Carl Helm, Arthur Jensen, Paul Diederich, Ross Quillian and Daniel Bobrow, Marvin Minsky, Arthur Anger, Bruce Ressler, John Moyne and David Loveman, Leslie McLean, William Cooley, John Carroll, Larry Wightman, Stanley Petrick and Jay Keyser. William McColly early provided us with the original data and worthwhile ideas. And Julian C. Stanley has served as a constant source of encouragement and inspiration.

Those readers seeking a shorter and more general introduction to this project are directed to the various publications by the workers, listed in the References. For a summary of this writing, they may wish to read the first section of Chapter IX of this report.

Ellis B. Page
Storrs, Connecticut

CHAPTER I

INTRODUCTION

When this research was proposed, the time surely seemed ripe for a much expanded study of computer analysis of student essays. In recent years rapid strides had been made in computer hardware technology, in the programming of language-data processing, and in linguistic analysis. More was known than formerly about the simulation of cognitive products and related fields. Many of the building blocks, therefore, appeared to be in place or nearly so. What remained was to thrust forward into the applied and basic problems of essay analysis and grading.

This study, therefore, aimed at advancing the knowledge of automatic essay analysis as far as theory, practice, and facilities would permit within the rather narrow span of time permitted. And this report will explain what was designed, attempted, and accomplished during this study period in this very new and potentially important field of research. It will also set forth current understandings about the most profitable avenues for further research.

And this first chapter will explain the background for the problem, both practical and theoretical, as well as the specific nature of the research attempted.

(A) The practical background. The practical problems of "objective" grading have long troubled education and the field of psychometrics generally. A single judgment of an essay by a single human judge is slow, extremely unreliable, and of uncertain status. When sufficient training is used, and a sufficient number of judgments establish a decent reliability, essay grading becomes prohibitively expensive. Psychometricians have therefore settled for multiple-choice items. These have the virtues of wide sampling, since more questions may be asked within a given time period; of high reliability; and of defensible validity, since scores often correlate as highly with judgmental ratings as the ratings correlate with each other under ordinary conditions.

Nevertheless, educators are far from content with multiple-choice examinations as the ultimate criterion of achievement. They wish to call upon students for global, organized responses concerning large questions in substantive fields. They would like to ask, in testing self-expression, for direct demonstration of corrent and literate usage. They are often not satisfied by the statistical evidence because of inadequate understanding of this evidence, and their incomprehension poses a problem for the psychometrician. More importantly, two objections to multiple-choice testing cannot be refuted comfortably at the present time: (1) One virtue of any test is the practice which the testing session gives the student. And it seems clear that the practice experiences of the student in taking an essay test are not precisely the same as in taking a multiple-choice test. (2) Another virtue of any test is the type of study which its anticipation motivates in the student before the test is administered. Many persons believe that students study differently for an essay test than for a multiple-choice test, differently for "recall" items than for "recognition" items. Clearer evidence on these two objections is needed, but their present status supports the desirability of finding some fast, reliable, inexpensive, and "objective" system of essay grading.

In English instruction especially, we have an example of a troubled field for essay analysis. Many believe that students need far more practice in writing essays in elementary and high school years. Yet writing without feedback seems generally pointless, and is surely objected to by the students concerned. And the feedback is very difficult to systematize. To do the ideal job in essay analysis, the high school English teacher would have to spend tremendous amounts of time out of class. Equalizing the load of the English teacher with his colleagues in other subjects is an unsolved problem. "Lay readers" are tried on an experimental basis in a number of schools, but these are an additional expense, are relatively untrained, and pose some large problems of coordination and aptness of judgment. Furthermore, the supply of qualified and interested English teachers has always been too limited. It is hoped that some way might be found to employ more broadly the talents of the few, so that individual judgment and correction of essays might be disseminated in the same way as lectures may be filmed or exercises may be printed in textbooks. A proper program for correction of essays would therefore be an attempt to amplify the effectiveness of the more intelligent and talented of graders and correcters. This study therefore aimed at the type of essay analysis most characteristic of English classes.

The input question. To solve any of these general practical problems would of course require practical input and output. At present, no computer does an adequate job of reading ordinary printing or typing, let along ordinary handwriting, into correct card images for further data processing and analysis. Yet rapid strides are being made in such recognition, and one may hope for resolution of input problems before the judgmental problems are completely satisfied. The computerized optical reading of standard typescript may be only a very few years away. Or, for that matter, the gradual replacement of much of student handwriting in the schools by inexpensive and noiseless character printers (perhaps related to the present Stenotype machines) seems a plausible and perhaps early development. But even with the present necessity of key-punching IBM cards from student copy, practical input for computer grading is not wholly out of the question. For example, the cost of such key-punching ranges below $2.00 per essay. Such an input cost, while out of the question for daily classroom routine, would not be unreasonable for an occasional master analysis, serving as a basic for extensive descriptive or prescriptive reporting, for screening or placement, or for certain other types of evaluation or guidance activity. Indeed, present objective-test batteries often cost much more than that. For the purposes of this study, however, it was assumed that input had been transformed into punched cards or card images, and concentration was on the correction and evaluation problems themselves.

(B) <u>The theoretical background</u>. The rather momentous practical consequences of computerized essay grading will be some years away. Before these are felt, there were theoretical questions important to the study, and there are theoretical answers which may be furnished by the study. These were psychological and linguistic in nature. Psychologically, for example, what roles do the actual various prose characteristics play in the cognitive and effective rating processes? Actual manipulation of prose characteristics is not anticipated in the present design, and therefore direct causal relationships will not be inferrable, but some important implications for these processes may turn psychological experimentation into some fruitful channels.

As a linguistic example, there is the additional understanding which may be gained of the nature of prose description. As Francis (1958) has pointed out, there are several kinds of "grammar": among them the prescriptive grammar, or "etiquette," of the schools, and the descriptive grammar characteristic of modern linguistics. (Also see "What Grammar?" by Gleason, 1964). It may be noted that computer analysis of this proposed kind produces still another sort: a set of descriptions resulting from the computer's own peculiar limitations and abilities. A list of prepositions may be employed, for example, and any match with this list may cause a counter to be incremented. In such a program, some words will be counted which the competent human judge would classify in other ways: as adverb, subordinating conjunction, coordinating conjunction, etc. Yet from this NPREP count may result a description which would be impractical for human judgment, which is 100% reliable within the essay, which probably has high reliability across essays of the student, and which may be useful in predicting the qualitative human judgments of the essays.

Furthermore, it was intended to use certain extant computer analyzers from other researches, and this was done. These are efforts to perform linguistic analysis within the sentence, and they are inevitably limited in accuracy. The limitation in accuracy need not be a handicap, however, in terms of useful theoretical and practical description.

The important point here is that the computer may provide new measurements of language usage and these will have inevitable importance for theory building and basic discovery. These measurements do not presently carry heavy theoretical freight, only because they have not been observable within the traditional technology. (See later discussion on this point.)

More will be said in the final chapter about theoretical outlooks for such research. It is enough here to note that both practical and theoretical interests motivated the present study.

Related Research

The field of essay evaluation by computer represents a new focus within the (also new) field of computational linguistics, just as it represents a new and divergent speciality within educational measurement and educational technology. Like all promising new areas of scholarly investigation, however, it must draw heavily upon some combination of background disciplines not ordinarily considered together. This section on related research will consider some materials from these background disciplines.

(a) Background disciplines

(1) Psychometrics is a basic discipline within which any system of evaluation must be justified. The discipline already has achieved many technical skills (assessment of various forms of reliability and validity) necessary to proceeding with the study at hand. Some of the particular psychometric problems in content analysis are discussed in work by Dexter Dunphy (in Stone, 1966). Important background work dealing with the reliability of essay grading by human judges has been done by Diederich, French, and Carlton (1961), by Myers, McConville, and Coffman (1963), and by McColly and Remsted (1963), to name only three outstanding recent examples. In recent years essay testing has apparently seemed so unprofitable to psychometricians that it has been almost wholly neglected. For example, the index of a recent Review of Educational Research about testing had only one item referring to essay testing and it is negative: "problems of unreliability in grading" (Merwin and Gardner, 1962).

(2) Linguistics has potentially very high relevance to computer analysis of essay examinations. Important lines of study have of course emerged from the "generative grammar" thinking of Chomsky (1957) and others (e.g., Miller, 1962; Postal, 1964). The implications of some of these more scientific approaches to linguistics for a broader psychology of language have been recognized by Carroll (1964) and others.

Of course, the particular newer field of this discipline known as computational linguistics is more intimately related to the present phases of this work. And this field in turn has a large overlap with the field of list-processing (see below), and of information retrieval. Many of the most effective workers in these fields come not directly from linguistics training, but from mathematics, psychology, and computer science.

(3) Curriculum. Curriculum, in all fields using essay examinations, is a concern of central relevance to the study. This is especially true of language arts education, where there are tensions (Gleason, 1964) between the modern descriptive linguist and the traditional "prescriptive" grammarian (such as Hodges, 1951, or Warriner, 1951), and what should be taught in composition is by no means certain (Marksheffel, 1964). Eventually, decisions must be made about the "right" approaches for any computerized master analysis. But for a problem of optimization of simulation of human ratings, hypotheses from both camps appear useful, and may be empirically checked against the criterion. And some interesting light has been cast on certain questions of the "etiquette" grammar by work already done with this project.

Although the language arts curriculum is especially important, it is by no means unique. Within the present research design, the study should produce some interesting information for curriculum within other key disciplines (see the procedures), especially regarding the importance of special vocabulary.

(4) Automatic language-data processing has been well described by a number of writers (Green, 1963, ch. 13; Borko, 1962, pp. 336-423), but one of the best general accounts is by Garvin and others (1963). In general, there appear two major methods which are possible: one is the content-analytic approach, like that used in the "General Inquirer", (Stone, et al, 1966) and is more a "statistical" method; the other is more oriented to syntactic and semantic relationships, as are necessary to the machine-translation studies underway, and may be considered a more "linguistic" method. Both appear promising for essay grading. Of particular potential help appear to be certain grammatical-classification computer programs already devised: a part-of-speech decider which is about 95% accurate (Stolz, Tannenbaum, and Carstensen, 1965?), and a dependency classifier (Klein and Simmons, 1963), which lists the various different structures possible for a given sentence. Especially significant are two systems already tried with small subsamples of our data, programs by Kuno (1964), and by John Moyne of the IBM Boston Programming Center.

(5) Statistical methodology is like psychometrics in having a great body of well-developed doctrine and practice which may be brought to bear on the present problem. An optimization solution may be sought with some standard statistical techniques such as multiple regression (e.g., Cooley and Lohnes, 1962); or in some sequential, decision-making form, such as an operations flow with a series of choice points (cf. Simon, 1964); or in some combination of the two. The verbal protocols of human raters might lead eventually to some appropriate combination.

(6) Computer technology is very important in both hardware and programming. Advances in machine design, especially in larger memories and reduced costs, will make feasible the more complex grading programs at more economical levels. But present equipment is adequate for extensive exploration of the problem.

Great strides have also been taken in designing software suitable for language processing. List-processing third-level computer languages are especially appropriate, and at least three have been written which are extensions of the FORTRAN framework: IPL-V, SLIP (Weizenbaum, 1963), and DYSTAL (Sakoda, 1964). Another important list processing language is COMIT (Yngve, 1962a, 1962b), designed for such work as machine translation. A modification of COMIT has been made by Stone (1964) and his associates for the "General Inquirer" system at Harvard. (After considerable investigation of computer languages, the present programming was, except for minor subroutines, entirely done in FORTRAN IV. This decision makes possible maximum versatility, availability of programmers, and dissemination of programs.) Two new developments in software promise increased ease of programming within AEC. One of these is STUFF (Puckett, 1966), which provides for string-manipulating functions embedded in FORTRAN IV. The other is in PL/I list-processing (Lawson, 1967), which is promised in an early implementation of the IBM 360 series (which has been installed at the University of Connecticut in August,1967).

One of the present lines of work in the field is that of the General Inquirer (Stone and Hunt, 1963; Stone, et al 1966; Ellis, 1964; Ogilvie, Dunphy, et al, 1962). For certain purposes, a short dictionary of under 4,000 root words has accounted for 90-98% of the ordinary written languages analyzed by General Inquirer (Dexter Dunphy and Marshall Smith, personal conference with the investigators December 22 in Cambridge, Mass.). Dictionary lookup procedures are crucial to language-processing, and recent developments of IBM research promise speeds of dictionary reference up to 10,000 words per minute (Philip Stone, 1964). As mentioned elsewhere in our proposal, studies by Simmons and others at System Development Corporation, by Stolz and others at Wisconsin, and by Kuno at Harvard have made progress in relevant software development.

Still another major line of automatic language-processing appears to be the movement toward what may best be called "computational humanism," especially concerned with data processing to solve the kinds of problems (concordances, attribution, influence, style) usually associated with literary scholarship. This movement is rapidly gathering momentum with conferences, workshops and institutes, and a beginning literature, such as the recent book by Bowles (1967), or the emerging journal, Computer Studies in the Humanities and Verbal Behavior, now being printed by Mouton Press, of the Hague.

These six fields, then, contribute to the background expertise which is producing a new and potentially useful sub-discipline within educational research. The analysis of essays by computer is seen to be based upon a number of other disciplines, some going back into the nineteenth century, but others part of the general growth of behavioral science and computer technology within the last several decades.

Objectives of the Research

In general, the objectives of the present study did not lend themselves to the clear, Fisherian, "classical" experimental designs, because not all operations could be foreseen. It did, however, permit clear procedures of dynamic development and exploration at each stage of the study, and clear verification of accomplishment at the end.

Properly understood, these characteristics are not handicaps, but symptoms of large research scale. In a recent paper, Baker (1965) pointed out that the larger and more exploratory research project "must be inherently dynamic and possess the ability to change its internal structure without sacrificing the rigor of the design" (p. 15). And another writer (Doyle, 1965) has recently stated that as a study approaches the "basic research end of the spectrum, it becomes more and more imperative to be free to alter the plan. Indeed, in basic research altering the plan ought to be a state of mind." With the present work, it would be mistaken and even misleading to commit the investigation prematurely to too narrow a path.

In general terms, the objectives of the present study were as follows:

(1) To identify important characteristics of student prose which are analyzable through specially devised computer programs. These characteristics were to be aimed especially at predicting human judgments of content, organization, style, mechanics, and overall quality.

(2) To develop computer programs for measurement of these qualities, or variables related to them, as they occur in school essays.

(3) To analyze the computer-generated objective data in relation to subjective measures of the essay dimensions, in order to improve the differential accuracy of evaluating such essay dimensions.

(4) To develop through this procedure greater understanding of the human rating process, as applied to objectively describable prose characteristics.

(5) To study those aspects of essay description which appear most promising for useful feedback to the teachers and students. In other words, to begin exploration of the feasibility of computer commentary about student essays.

(6) To set forth larger strategies for the most promising future exploration of computer grading of essays.

This report tells about the pursuit of these objectives, in the following chapters.

CHAPTER II

THE BASIC DESIGN

Some fundamental strategies of investigation were designed early in 1965, and employed in the first data runs of Project Essay Grade (PEG I), financed primarily by the College Entrance Examination Board. But that study was intimately involved with the present one, and merged into it, and completely separate reporting of research done under the two sources of support would do some injustice to this continuity. Furthermore, although there has been much reporting of all of this work in professional publications, at scientific meetings, and in more popular news media, there has not been a disseminable technical report of any of it. Thus this report will at least touch upon all of the work to date.

Rationale

We should begin with a general rationale concerning the computer grading of essays. This presentation seems necessary for two reasons: (1) The computer analysis of essays seems to some a radical proposal, and is not treated elsewhere in psychometric literature. (2) The investigators intend the present project to open a larger exploration of such measurement and feedback, with possibilities not at all limited to the present work.

In general, then, there appear to be at least two dimensions of the problem of essay grading, with two general approaches in each dimension. In the first place, there is the content vs. style dimension. Are we interested in *what* the student says (e.g., about the discovery of America by Columbus), or in the *way* he says it (e.g., his use of punctuation)? We all know that these categories are not mutually exclusive, but they are useful concepts for our first orientation (Page, 1966).

In the second place, there is the dimension of rating simulation vs. master analysis. Are we interested in an actuarial approximation of the ratings of human judges (e.g., in certain words statistically associated with high ratings, even though not themselves regarded as an index of correct expression)? If so, we are essentially interested in rating simulation. Or are we interested in the computer doing a "reading" of language and performing a kind of informed and rational "judgment"? If so, we are speaking of the computer as master analyst, and of creating a kind of "artificial intelligence." These two dimensions are pictured in Figure II-1.

	I Content	II Style
A. Rating Simulation	I-A	II-A
B. Master Analysis	I-B	II-B

Figure II-1

Possible Dimensions of Essay Grading

Clearly the columns of Figure II-1 are not going to remain unrelated to each other, since in some ways content and style are inseparable. And the column headings given are not completely satisfactory. Spelling, for example, is a consideration in Column II, yet "style" does not appear a satisfying rubric for the marking of spelling errors.

Similarly, Rows A and B will not remain unrelated either. As the investigation of simulation discovers variables which are, empirically, more and more accurately correlated with human ratings, the analysis will become more profound and will grow closed to the "meaning" analysis eventually necessary in Row B. The top row, then, suggests the "actuarial approximation" to judging the essay, and the bottom row represents the "master analysis" of the essay itself. These rows represent matters of computer strategy and objectives.

These rows need further explanation, because they are very near the heart of the problem, hence are crucial to understanding our progress to date in Project Essay Grade. What we have taken as our first goal is the imitation, or simulation, of groups of expert judges. How we reach this goal of successful imitation is not the central question, so long as it is reached, and so long as we can actually match or surpass the human judge in accuracy and in usefulness. In attacking the problem in this way we are clearly not doing a "master analysis" or generating measures of what the true characteristics of the essays are, as ordinarily discussed by human raters. Rather, we are content to settle for the correlates of these true characteristics.

To express this important distinction, we have been forced to coin two words: trin and prox. A trin is the intrinsic variable of real interest to us. For example, we may be interested in a student's "aptness of word choice," or "diction." A prox, on the other hand, is some variable which it is hoped will approximate the variable of true interest. For example, the student with better diction will probably be the student who uses a less common vocabulary. At present, the computer cannot measure directly the semantic aptness of expression in context, or "diction." But it can discover the proportion of words not on a common word list, and this proportion may be a prox for the trin of diction.

Or another illustration: We may be interested in the complexity of a student's sentences, in the branching or dependency structures which he has the maturity to employ. Such sentence complexity wo'ld, therefore, be a trin. But the sentence-parsing progra. s for computers which exist now are not completely satisfactory for our purposes. We might therefore hypothesize that the proportion of prepositions, or of subordinating conjunctions, constitute a prox for such complexity. And we might therefore employ this proportion, too, in our computer analysis.

One more essential, and the basic strategy of our first essay grading project may be understood: We have begun by saying that the basic evaluation of overall essay quality must be human. But which human? If only one expert English teacher grades an essay, we know that the judgment will not be very dependable. We know that other judges will reach a somewhat different conclusion, and even the same judge, if he were grading it again, would probably shift his evaluation. The typical inter-judge agreement is represented by a correlation coefficient of only about .50. On the other hand, when a group of independent experts have graded an essay, and when these grades are averaged, this average has a rapidly improving dependability. When four judges, for example, grade an essay independently, their average judgment will correlate with the average of four other judges about .80. So it is possible to get reliable human judgment of essay quality. But it is extremely, prohibitively expensive and time-consuming when applied to any large-scale testing.

However, getting a reliable human judgment is not too expensive for a sample of essays. If we can find a way to imitate, then, what the expert human judges do with this sample, and if we apply this strategy to a computer program for a huge number of other essays, we capture high quality of judgment at low cost. And the techniques used to analyze the judgment and reproduce it are essentially those already so well developed in standard prediction problems.

The strategy, then, is very general indeed: if the computer may be programmed to simulate some sample, the resulting algorithm may be employed on arbitrarily large numbers of essays drawn from the same population as the sample. The validity of any evaluation and analysis will then depend on basic conditions which are already very familiar, from measurement work, to the psychometrician: on the number of judges used to establish criterion evaluations; on their quality; on the "set" of the judges; on the number of essays evaluated; on the nature of the essay sampling; on the frequency and consistency of the proxes; and so on. And powerful, well-understood statistical tools may be brought to bear on the simulation.

One technique for such simulation, where the appropriate weighting of each prox is unknown beforehand, would be the familiar multiple regression, in which one cri- rion variable (in this case the human judgment, or trin) may be optimally predicted by a discovered weighting of a number of predictors (in this case, the computer proxes). And indeed, this general tool of multiple regression, implemented by appropriate computer programs, has proved very powerful for essay grading, both in the initial strategies and in the later ones.

To summarize the general design, then: (1) Essays to be evaluated must (at present) be key punched for computer input. (2) These essays must be independently evaluated by human judges (of any desired characteristics), on various traits (depending on the research hypotheses). (3) Hypotheses must be generated by other human experts, concerning the programming of appropriate proxes for evaluation. (4) These hypotheses, depending on convenience and promise, must be programmed into the computer analysis. (5) The machine-readable essays are passed through the computer, and the proxes recorded for each essay. (6) These proxes are then optimized for the best possible prediction of the pooled human judgments.

The flexibility of the general design is clear. It allows for any appropriate selection of judges, any selection of proxes, of traits to be predicted, of essays, etc. Thus, this design has a great capacity for repeated use as our knowledge of essay grading broadens and deepens, and as its concerns expand to include all parts of the universe of Figure II-1.

In this study, the attention first focused on simulation of ratings of overall quality of style. Then the concentration shifted to ratings of various essay characteristics (content, organization, style, mechanics, and creativity). A variety of subproblems were considered, and hypotheses tested, and phrase-recognition procedures were implemented. And currently, attention is expanding to include subject-matter knowledge exhibited, and more intensive linguistic strategies. But the basic design is easily adapted to these and other shifts of focus, as research interests become more sophisticated, and exhibit greater breadth and depth. Indeed, even with the advanced strategies projected in the final chapter of this report, it is difficult to imagine a time when such actuarial strategies will not constitute an important part of some final decision process.

CHAPTER III

THE INITIAL PROXES

This chapter will describe more of the fundamental thinking to date about computer analysis of essays at the University of Connecticut. First this report will consider the 1965 work, which predicted judgments of the overall writing quality of a set of essays, and second the later expanded work, predicting a more complete profile of judgments on a number of essay characteristics or traits. This particular chapter will be concerned with the sampling, procedures, proxes, and programs devised for such analysis.

Sampling. The basic research design has been described in Chapter II. Since there was great flexibility permitted in selection of essays, and since the investigators were eager to explore the parameters of this field, a search was conducted for essays which would have certain desired characteristics. What seemed desirable were essays which (1) were already written under carefully described circumstances; (2) had ratings by multiple human experts already assigned, independently of one another; (3) were drawn from a student population heterogeneous enough to furnish a reasonable reliability for rating sums; (4) were long enough to furnish stable measurements of at least some prose characteristics; (5) were multiple for each student, so that some estimate could be made of test-retest reliability; (6) were general enough so that findings might have fairly wide applicability; (7) were accompanied by correlative information about the students; (8) were representative of a random sample of the target student population; (9) were large in number.

A sample of essays fulfilling most of these requirements was obtained in 1965 through William McColly, then of State University of New York, Oswego. For an earlier experiment in composition teaching, McColly and Remstad (1963) had arranged for English classes at Wisconsin High School (Madison) to write four essays, on four different topics, about one month apart. These had been indeed (1) written under carefully described circumstances; (2) given four independent ratings for "overall writing quality"; (3) drawn from a heterogeneous student population, representing grades eight through twelve, with an average IQ of about 114; (4) of an average length of over 300 words; (5) four in number for each student; (6) written on rather common themes, such as whether the "best things in life were really free", or whether "anger" could have good uses; and (7) accompanied by fairly extensive information about the student writers. Since they were from one (rather atypical) high school, they could not be said to represent a random sample from the secondary population of the United States. On the other hand, for such an exploratory research, the proposed experimental analyses were so broad that subtle interactions with ability levels, or with other levels of student population, were believed of small initial concern. Finally, the number of the essays was substantial, with well over 250 essays for each of the four writing sessions. For multivariate analysis especially, large numbers of cases are very important.

The question of interjudge reliability is of great importance, since any optimization technique, such as multiple regression, must have a decently reliable criterion if it is to produce any nonrandom results. The overall ratings assigned by the Wisconsin judges had an average interperson agreement of about .5, and an analysis-of-variance reliability for four such judgments pooled of around .83 (McColly and Remstad, 1963, p.49). This high a reliability would give the sums (or averages) a sufficient stability for use as a criterion.

Hypotheses and proxes. Having defined the criterion and established a suitable sample, the next important task was to determine what hypotheses were appropriate, i.e., which of the available hypotheses could be shaped into suitable algorithms to provide proxes for the multiple regression. Clearly, it would have been ideal if we could have incorporated into a massive computer program nearly the whole of standard texts on usage and rhetoric, such as the Harbrace Handbook (Hodges, 1951). That is, in one sense, still the target of such work, but no one dreamed that anything approaching such a goal could be implemented into the study at such an early time. The problems were not simply economic and logistic. More importantly, they stemmed from fundamental uncertainty about the nature of language and of the human reading process. The present status of such work will be considered under suggested future strategies. Here shall be discussed the sort of thinking generated in conferences of consultants (Daigon, 1966).

The agreement between independent raters of the essays will indicate the degree to which the essays themselves (rather than the independent personalities, moods, biases, etc., of the judges) influenced the ratings. That is, the inter-rater agreement is a function of the physical influence of the word patterns of the essays. In principle, therefore, the computer is limited in its simulation of the group judgment not by any spiritual nature of the essay itself, but only by the extent to which the computer program can be designed to reflect the group responses (Page, 1967b).

These group responses may be presumed to be related to certain intrinsic characteristics of prose. These intrinsic characteristics may deal with mechanics, with organization, with diction, etc. They are described in detail in prescriptive grammars, and elsewhere, and may be further elaborated by the project's investigators and consultants. On the other hand, some characteristics of ultimate interest, some trins, may be unmeasurable with present knowledge and technology, and some possible approximation to them may be studied, in the hope that these second-order variables will be correlated with the trins.

As one example, spelling may be considered a trin, or almost so. The simplest effective strategy for analysis of spelling with available computer technology was to use a list of misspellings. A list of several thousand common spelling errors in their misspelled forms (e.g., Gates, 1937, with later supplement) will, consultants agreed, possibly account for many misspellings in high school papers. Each word in each essay may be looked up in such a computer-stored list, therefore, and a student's "misspelling score" augmented by one point whenever such a word is encountered for the first time. Not all student misspellings will be discovered by this method, but scores so generated would be correlated with the "true" spelling scores as might be discovered by human examiners, and any given misspelling is a trin. There are other available trins. Ungrammatical combinations of words, examples of generally poor diction, and other solecisms may be similarly discovered and tabulated from comparison with such lists, and may also be considered trins, considered individually.

On the other hand, what of the "less mechanical" questions of content, organization, thought pattern? Let us consider an example of a prox: The Harbrace College Handbook (Hodges, 1951) contains a chapter on "the paragraph." Surely the judgment of paragraph organization is one of the loftier goals to which the project may aspire, and a fully satisfactory simulation may be some good time away. But consider certain rules given by Hodges for the paragraph. His Rule 31b is:

> Give coherence to the paragraph by so interlinking the sentences that the thought may flow smoothly from one sentence to the next. (p. 330)

This rule is of course too general to afford much help. But Hodges has given more prescriptive help in the five sub-rules [each provided with examples not reprinted here]:

(1) Arrange the sentences of the paragraph in a clear, logical order.

(2) Link sentences by means of pronouns referring to antecedents in the preceding sentences.

(3) Link sentences by repeating words or ideas used in the preceding sentences.

(4) Link sentences by using such transition expressions as the following:

ADDITION moreover, further, furthermore, besides, and, and then, likewise, also, nor, too, again, in addition, equally important, next, first, secondly, thirdly, *etc*., finally, last, lastly

[etc., through other longer lists]

(5) Link sentences by means of parallel structure --that is, by repetition of the sentence pattern. (pp. 330-335)

These rules suggested some good researchable hypotheses. .. Number (4), with its extensive list of words believed appropriate to link ideas in different ways, was the most convenient, and was researchable through a straight dictionary-lookup procedure like that used for spelling. The question is then to what degree such words may be a prox for the trin of paragraph organization. Similarly, Number (3) may be researchable, if the repetition of *words* is alone researched. The repetition of *ideas* would clearly depend on a dictionary or thesaurus beyond the scope of the immediate project. For Number (2), a prox might be the number or proportion of such pronouns occurring after the first sentence in any paragraph. (The complicated questions of pronoun reference again depend on distant developments in semantic and syntactic analysis.) Hodges' other rules may perhaps be approximated rather remotely, but argue for developing or adapting a syntactic sentence analyzer.

Another example of a trin was word fluency. This variable was clearly difficult to measure mechanically, since it would often depend upon semantic understandings, and these were generally beyond the scope of available technology. Nevertheless, possible proxes suggested themselves. Lists of "common words" exist (Lorge, 1959). The words of essay text may be looked up in such lists and, where unlisted, scored appropriately. The ratio of such unlisted words to total number of words may be included in the multivariate analysis to determine whether it aids in predicting evaluative rating. Or another approach, closer to a "content" analysis, would be to check for the presence of certain words suggested by dictionary or thesaurus as synonyms or near-synonyms of some thematic words. And extensive work of this kind is currently underway in a new phase of the research.

In short, the hypotheses for the trins underlying the human ratings were very numerous, and preliminary thinking of this sort, both initially and through the following two years of work, occupied a fair share of the time of consulting experts. As always with multivariate research, it would be far too cumbersome to recount the entire chain of thinking leading to each specific prox employed, yet some explanation will be included in the next section. The most obvious and general hypothesis for all trins was that the papers receiving better human marks would tend to be written in a style more conformable with the standard textbooks.

Hypotheses and proxes. The first 30 proxes which we settled upon grew out of several considerations: (1) We would first decide which trins were ideally measurable; but as we have seen, such a list included almost the entire handbook of usage, with most points defined very intuitively. (2) We would then decide what short-cuts might be taken to an approximation of such trins; where these were easily manageable, they would be programmed into the analysis. (3) We would furthermore have, from the nature of our text analysis, a number of variables which would be fortuitously and easily come by; and these might be examined routinely for possible assistance in prediction.

Ordinarily, as almost all methodologists believe (e.g., Tatsuoka and Tiedeman, 1963), research should be primarily theory-oriented, i.e., directed by hypothesis and associated deduction. Yet multivariate analysis does not really lend itself to complete explication and text of each separate hypothesis, and in general prediction research would be unnecessarily and artificially restrained if it were not permitted use of any convenient predictors, regardless of the vagueness of rationale for their inclusion. There were in this study a fair number of what might be called, therefore, "proxes of opportunity." Some data about each of the initial proxes will be reported later. Here they will be listed, and briefly explained.

1. Title present or absent. It was early noticed that some students did write a title, and some did not. It was guessed, provided there were a fair division on this point, that the better students would be somewhat more apt to compose titles; and there would be therefore an expectable positive correlation with human ratings.

2. The average sentence length is a variable of considerable interest. If a sentence is defined the way the student writer defines it (that is, as a string of words between non-abbreviating periods), then there is not

much evidence to expect more than a slight correlation with quality. Kellogg Hunt, for instance (1966), has shown that mean sentence length remains fairly constant with advancing school age. On the other hand, it might be supposed that a combination of sentence length and dependency relations would be reasonably important; that sentence length without such internal dependencies might be a sign of the poor writer, the run-on style; but that sentence length with such dependencies might be a sign of greater language maturity.

3. The number of paragraphs will often be very small for a really immature writer, just as other forms of linguistic markers and conveniences will also be underutilized. Thus it was predicted that frequency of paragraphs would be positively correlated with writing quality.

4. Subject-verb openings are the sentence beginnings where the subject phrase is apparently first. Without a parsing program, this variable was only approximated, and it was done so on the assumption that the first word would in the majority of cases be adequate for decision. Any pronoun, article, abstract noun, etc., will typically signal a subject opening, whereas an adverb, subordinating conjunction, etc., will typically signal a left-branching sentence. An essay's score on this variable, then, would be represented as a ratio of subject openings to total number of sentences. A common youthful failing is a stodgy, mechanical style without variation, while the sign of the more mature writer is a variety of sentence structures, depending on the purpose of the sentence. Therefore the prediction was that the subject-verb proportion would be negatively associated with writing quality.

5. Length of essay in words is surely a characteristic associated with advancing maturity and skill; and it is a commonplace correlative of high ratings from human judges.

Here the prediction was that essay length would help in the prediction of the mark received, and would be positively correlated with writing quality.

6. The frequency of parentheses might be supposed characteristic, in a high school sample, of writing fluency. Among poor writers, many of these common tools do not seem to be a part of the available repertory, and it might therefore be predicted for parentheses, as for other marks of punctuation, that they would be positively correlated with writing quality. (Here and for similar subsequent counts, the frequency should be taken to mean a ratio of the item to the appropriate total of the essay. In this case, the number of words is used as the control for length. Otherwise, length of essay would be a hidden, contaminating factor in most of the proxes.)

7. Apostrophes are in a somewhat different category. While it is plainly more correct to write DON'T than DONT, it is somewhat better usage, or at least more formal usage, to write DO NOT. Frequent apostrophes might be supposed to mark a rather informal or casual style, and it might be supposed that informality is on the whole negatively regarded in a set theme assignment. On balance, therefore, apostrophes were predicted to correlate negatively with writing quality.

8. The frequency of commas might be the most reliable measure of the student's repertory of punctuation facilities, since commas are more common than any other mark. It was predicted, then, that comma frequency would be positively correlated with quality in a high school setting.

9. The frequency of periods is not, like frequency of commas, a mark of writing fluency, since it may be evidence of short sentences, or of abbreviations. Neither of these would be considered an asset in such a formal assignment.

10. The frequency of underlined words was predicted to be slightly, but positively, correlated with writing quality, under the simple assumption which also governed parentheses and commas. Similar predictions were made for the following punctuations:

11. Dashes

12. Colons

13. Semicolons

14. Quotation marks

15. Exclamation marks

16. Question marks

and, out of order:

26. Hyphens

27. Slashes

17. Prepositions are an interesting frequency. In the first place, it was not possible to design an algorithm to be very sure about the accuracy of category. For the initial programs, a word was a "preposition" if it was found in a computer-stored dictionary of prepositions, though to the human expert it might be serving as an adverb or subordinating conjunction, etc. Prepositions are common words, of course, yet it was predicted that they would be positively associated with writing quality, simply because their frequency would imply dependency substructures within the sentence. When sentence length is held constant, as was noted for #2 above, one might suppose that preposition frequency would vary positively with quality.

18. Connective words, such as nevertheless, however, and also, were assumed to characterize language marked by complexity of relationship, and thus were hypothesized to correlate positively with writing quality.

19. Spelling errors are of course the most obvious and objective characteristic of writing which is poor mechanically. In this test, no attention could be given to the errors which are simply misplaced homophones (such as THEIR and THERE), nor to other errors which were guessed low in frequency. Rather, the list consisted of some of the commonest misspellings which are wrong in any context (e.g., THIER, BELEIVE, DONT). And the assumed direction was that there would be a negative correlation between such occurrences and the human judgment of writing quality.

20. Relative pronouns are another set of words used by able writers to marshall and interrelate their thoughts. Therefore it was predicted that there would be a positive correlation between such words and essay quality.

21. Subordinating conjunctions were similarly expected to correlate positively with essay quality, for the same reasons as those above: that such words are important and relatively advanced tools for imbedding sentences and relating one thought to another.

22. The proportion of common words in an essay was determined by mechanically looking up each word in the Dale and Hall (1948) list of common words, and dividing the number of such occurrences by the total number of words in the essay. Setting aside misspellings (some of which would be caught by other dictionaries), we would expect that those essay words not on such a common list would probably be less frequent and more discriminating selections, and would usually represent better diction. Therefore we predicted a negative correlation between such common words and essay quality.

23. The occurrence of a sentence with a missing final period is very hard to find, with present computer programs. However, at the end of a paragraph, a missing period is

obviously easy to detect, and this mistake does occur among very immature or careless writers. It would be predicted that where such an error did occur, it would be negatively correlated with writing quality.

24. This item, declarative sentences type A, and the next item, treat an attempt to locate sentences where question marks are mistakenly omitted. Any sentence ending with a period was here taken to be a "declarative" sentence. Then the first word is examined to ascertain whether the sentence might be interrogative in syntax. If the sentence begins with any of the common question introducers, such as WHO, HOW, WHERE, etc., it is taken to be a "declarative sentence type B," meaning that there is a boolean conjunction of a possibly interrogative first word with a non-interrogative terminal punctuation. A "declarative sentence type A", then, is one in which there is no evidence for interrogative sentence either in the first word or in the terminal punctuation. From this algorithm, then, the sentence is consistently declarative, and may be better correlated with the criterion that would be the type B sentences.

25. For these "declarative sentences type B," therefore, one might predict, if anything, a negative correlation with quality.

26. - 27. Punctuation marks, already discussed above.

28. The average word length in letters might be predicted of considerable actuarial importance, because we know from Zipf's law that word length is correlated with word rarity, and word rarity may be presumed correlated with broader vocabulary and more accurate diction. Thus the predicted relationship with quality would be positive.

29. The standard deviation of word length might be presumed to be highly correlated with the length itself,

but it was thought that the additional information about dispersion might add to the total regression. This prox would also be predicted to correlate positively with the criterion.

30. The standard deviation of sentence length would not be presumed, necessarily, to correlate very closely with the length of sentence, since it is a common observation that many persons write consistently short sentences, or consistently long ones. What would appear ideal is mixture of long and short sentences, as appropriate to the context, and one would therefore predict a standard deviation of sentence length which would be positively associated with quality.

In summary, these initial proxes were justified partly on rational grounds, partly on common sense observations, and partly by expert opinion. As we shall see later on, most of the predictions were discovered to be in the right direction, though not all; and some were considerably less or more effective than we had foreseen.

<u>The Computer Program</u>. Having decided upon the basic proxes for the first studies, it was necessary to choose a programming language for their implementation. This is not a trivial decision, since the world of "natural-language" programming, as it is called, has been and is a rather chaotic one. For some large-scale researches, through the past years of programming for natural language analysis, efficiency has been extremely important, both for time and money considerations. Consequently, some of the most important work in language translation (see Oettinger, 1960), linguistic analysis (Garvin, 1963; Borko, 1967), content analysis (Stone et al, 1966), and information retrieval (Becker and Hays, 1963) has been programmed in symbolic languages close to the machine, such as FAP or MAP. And these low-level languages not only make changes difficult and buggy, but also are extremely difficult to move from

one machine configuration to another. Such programs are of little help to the new researcher in natural language work.

At the other extreme are high-level and sometimes quite abstract languages which have been used for frontier work in psychology, management science, linguistics, and artificial intelligence. Such languages are COMIT, IPL-V, DYSTAL, LISP, SNOBOL, and SLIP. These and others have been designed for list-processing, dynamic-storage applications, and often pay heavily in speed and convenience for the flexibility and elegance suitable to such applications. These were also surveyed rather extensively for any suitability for our system needs.

Ultimately, the choice of programming languages for such a purpose should be governed by these rather overlapping considerations: (1) Is it easy to program, and easy to modify? (2) Are the relevant programming skills already available in the research team? (3) Will the program in general outlive the rapid and inevitable machine changes across the years? (4) Will other researchers be able to adapt it easily? (5) Is it natural to our own systems tape? (6) Is it a mnemonic language, easy to comprehend?

In light of such considerations and after some false starts with COMIT, the investigators decided upon FORTRAN IV, for the following reasons: Our own computer installation at the University of Connecticut, was at that time a rather new IBM 7040, with extensive FORTRAN IV facilities as part of the regular system tape. FORTRAN was the most-widely used programming language in the computer world, with large numbers of available programmers. It furthermore promised to be available at almost all large computer centers for years to come. It is relatively machine-independent, with the exception of a few considerations of word-capacity and other matters.

Especially, FORTRAN seemed suitable because, when our problem was spelled out carefully, list-processing and dynamic storage were not yet necessary to anything we wished to accomplish. Such facilities are excellent conveniences for certain types of problems; but the better we came to understand our early needs, the more obvious it was that we needed the following:

(1) A way of organizing character strings into ordinary alphameric arrays, each row of such an array representing a recognizable "word", in the usual language sense. This organizer would also need to set aside punctuation marks and other non-words.

(2) A way of reading special dictionaries into immediate-access storage, for easy comparison with the words of the student text.

(3) A way of efficiently counting occurrences of such dictionary words, for any student sentence and any essay.

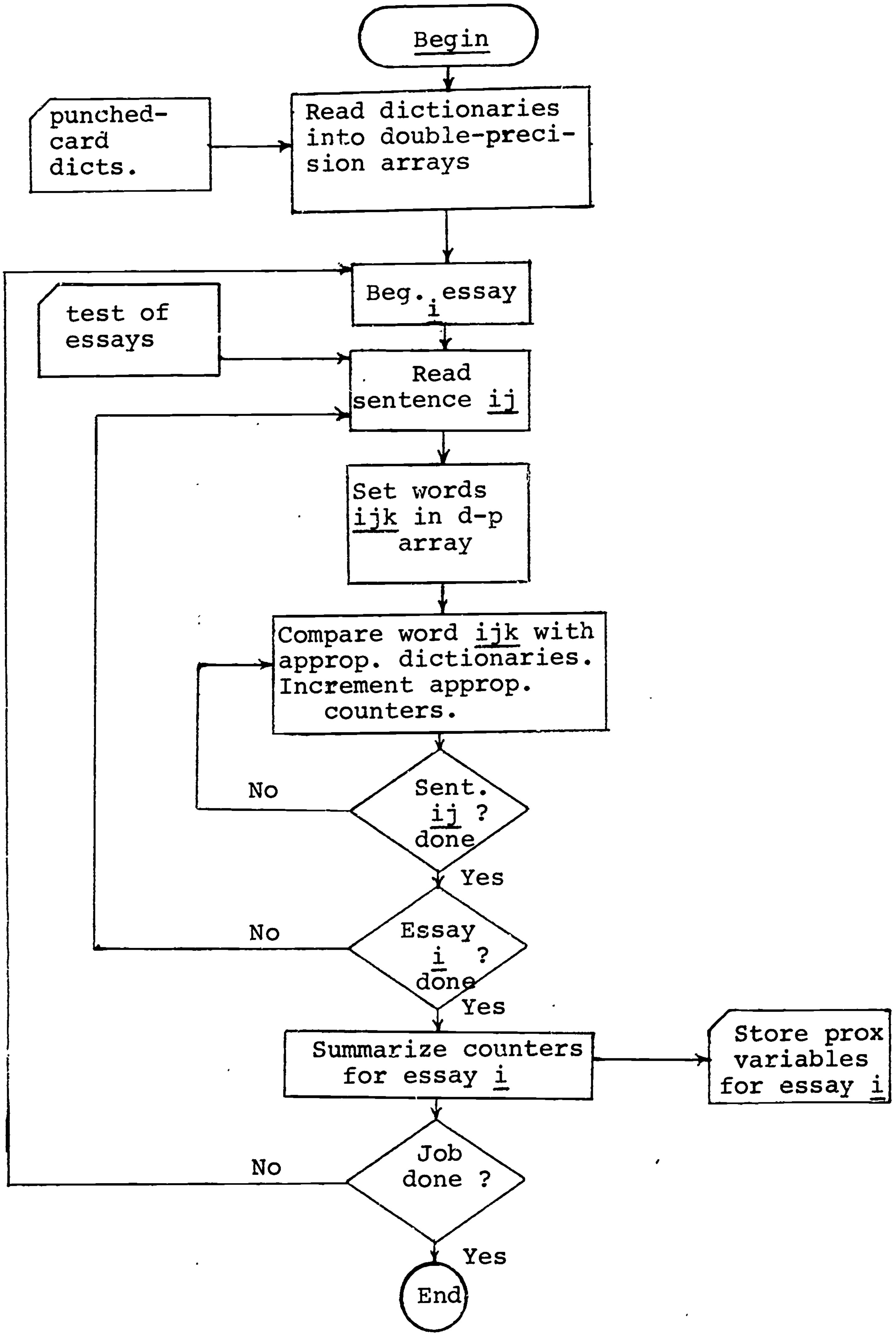
(4) A way of checking on various other, non-dictionary events in the student text.

(5) A way of summarizing the proxes for an essay.

These general goals are shown in only slightly more rigorous a way in Figure III-1, which is a flow chart of the first program outlines. Here it is seen that our dictionaries were input in punched cards, and were stored in core, in what are called double-precision arrays. For many readers, this requires some explanation. The core storage of the IBM 7040 was at that time limited to 32,000 computer registers, in which each register was limited to six characters of the alphabet, number system, punction set, etc. While the average English word (in running text) is between four and five letters in length, the average dictionary word (with small proportions of common words)

FIGURE III-1

GENERAL FLOW CHART FOR FIRST PROGRAM
PROJECT ESSAY GRADE
(ESSAY ANALYSIS)



will naturally be longer, and words will often be too long to fit within a six-character register.

For this reason use was made of a facility of FORTRAN programming called "double-precision" addressing, which permits a set of two such six-character words to be addressed as if it were one. This scheme permitted English words to be packed in up to 12 characters, but truncated any words longer than 12.

Since each word was originally read in from a punched card, 80 characters in length, the first problem of processing a sentence was to reorganize these characters into words. Such markers as spaces and punctuation permitted identification of such words, and these were then "packed" from the loose original array, which was organized with one character in each computer register, into the denser 12-character registers. Then these text words could be compared with the dictionary words by comparing the first six-character register of each word. If a match were made, the second six-character register was also examined, and if another match were found, a hit was recorded for the particular list examined. This method of "packing" such words, then, permitted two economies: a large economy of space, since 1000 English words could be contained in only 2000 computer registers; and a large economy of time, since a match of the first six letters could be made in just one arithmetic comparison of one cell with another.

As is shown in Figure III-1, the student essays were also input in punched cards, and the eventual proxes were output in punched cards as well. (Later systems are tape-based.)

This original FORTRAN IV program, as modified and used throughout the length of this present report, is listed with considerable comment in Appendix A. Since the accompanying documentation is fairly extensive, we shall not describe the program in any great detail here, although it is obviously one substantial product of the work. In general, however, the effort was to make a program that would be: (1) efficient, so that expenditure of time would not be too great; (2) modular, so that it might be easily understood, and altered as circumstances would require; (3) general, so that dictionaries, numbers, functions could be easily changed; and mnemonic, so that variable names would be reasonably easy to learn and remember.

An example of the modular and mnemonic nature of the program might be seen in the function which searches for a given text-word in any particular dictionary. This function is called INTABL, and appears in statements of the form:

```
IF (INTABL (WORD, PREP, 100)) GO TO 900
```

Here the argument WORD refers to the particular essay word to which the DO loop has brought us in our data processing. Let us say that such a word might be AFTER. The argument PREP refers to the sub-dictionary containing prepositions, which is stored in core, and may be quickly searched. And the argument 100 is the (maximum) length of that list of prepositions. The function INTABL causes the program to transfer to a subroutine, which makes a search in that list called PREP for the word (in this hypothetical case, for the word AFTER). If the word is found in the list of prepositions, then the function INTABL is "TRUE," and the command of the IF statement is followed. In the present case, this means a transfer to statement number 900. If the word AFTER had not been found in this subdictionary, then the operation would have moved to the next statement following the IF, whatever that might be.

The manner of the search may also be of some interest, since dictionary look-up is surely one of the principal operations in the program. In a completely random sequence an exhaustive search would have to be made through the list in question; this would be much too inefficient. Rather, some advantage may be taken of the alphabetical sequence, and of the fact that the order of the letters corresponds with the size of the binary numbers in which the letters are represented. This means that early letters (such as A) will be represented by low binary numbers (with many zeroes). This also means that a word may be easily compared with a given spot in the list, and it may be said whether that word matches it, or may be earlier in the list, or later. This is sometimes referred to as "equal to-less than-or greater than" comparison.

Such a comparison permits several techniques. The most obvious is to plod through the list until the point is reached where the word should be, alphabetically speaking. If it is not there, then the operation may be returned to the main program, with the value FALSE. This technique of using the alphabet in a straight linear search will, then, obviously save about half the search time for the word in question.

A more advanced search technique, however, is what is called a binary search. This operates by going at once to the middle of the list, and making the comparison at that point. If the word is earlier, then the first half of the list is divided, and a comparison is made with the list at that quarterpoint. The list keeps being narrowed by half each time a comparison is made, so that very soon the comparison is narrowed to a single word: if the text word does not match the list at this point, the operation returns to the main program with the value FALSE. Such a binary search obviously capitalizes on the great economy of the exponential number. And this is an economy which rises rapidly as the dictionary increases in size. The number of comparisons made will be about the logarithm base 2 of the number of words in the dictionary. That is, if D is dictionary size, and $D = 2^n$, then $\underline{n}$ is the number of comparisons required, in the usual case, to ascertain whether any word is present in the dictionary. Then if a dictionary is 16 words long, about four comparisons will locate it. This may not seem a large saving over the linear alphabetical search, when the time is added to compute the next comparison. But if a dictionary is 2,048 words long, a mere 11 comparisons will locate a word's proper space, and this binary search yields a great saving indeed.

Other lookup techniques, some even more economical in time, are discussed elsewhere (Hays, 1967, Chapter 5). Without such efficiencies as binary search, practical essay-grading would be prohibitively expensive. A number of other efficiencies were introduced into this program as well.

Preparation of the text. As we have said, eventual implementation will require some fairly direct input process from the student to the computer, at least for ordinary classroom use. For research purposes, we had these key-punched by clerks at the University of Connecticut, according to a fairly obvious format. Since at that time our key-punch machines had no upper-lower case differentiation, all typing was in capital letters. Also, the punctuation set was not complete, so that we employed the following conventions:

Name	Typewritten	Machine Convention
Period	.	.
Comma	,	,
Semicolon	;	.,
Colon	:	..
Exclamation	!	.X
Question Mark	?	.Q
Italics	____	(/)XXX
Dash	--	--
Apostrophe	'	@
Quote	"	*

Of course, these made no important difficulty in the programming, since two consecutive symbols are very easy to look for. In order to distinguish a period (abbreviation) from a period (end of sentence), we looked for two spaces after it, and took that to mean end of sentence. Similarly, a new paragraph was signalled by four blank spaces at the beginning of a new line.

Key-punching of these essays proceeded at about the speed expected of ordinary typing, although verifying might take somewhat longer than ordinary proof-reading. What was more time-consuming was that the clerks were under instruction to type the copy literatim, that is, including every last mistake of the student in spelling, punctuation, and word order. This took time, of course, because it would be contrary to the habits of a career devoted to eliminating such mistakes.

The most important aspect of the text preparation was that nothing was done to the text which was not required for it to be machine readable. In no case was any human coding of it done for any purpose of the subsequent research (for example to identify verbs, nouns, etc.). This means that the copy to be read by the computer was in almost every obtainable way just what the student himself would presumably have written, if he had known how to typewrite and had typed it himself on the key-punch.

Summary. This chapter has elaborated the sampling, hypotheses, proxes, programs, and procedures for the investigation of machine analysis of essays. And the principal program so fundamental to the work is found in Appendix A of this report. The next chapter will treat some results of importance from such analysis.

CHAPTER IV

PREDICTING OVERALL QUALITY

This chapter will describe some of the findings, and implications of the findings, from the attempt to predict the rating of overall quality of writing. This describes work done in 1965, 1966, and 1967, largely concerned with the data from the Wisconsin study, which has formed a focus for much of the research on style up to the present time.

Human ratings. As has been made clear from Chapter II, the principal strategy of the work has aimed at the simulation of human judgments, and these human judgments are therefore very important. The instructions used for the ratings in Wisconsin were described by McColly and Remstad (1963). They asked for ratings on "overall quality", and they had four independent judges for each essay, and four essays for each student subject. The individual judges were qualified, but their personal characteristics are not of much importance for our study, and the so-called "individual" ratings represent a kind of statistical artifact. That is, when essays are regarded as rows, and the judgments are represented in four columns, each of these columns is a kind of composite, since it may contain ratings from many of the judges used in the Wisconsin study. Each particular element in the column is a rating by one human judge, but the column as a whole may be the contribution of many such judges.

With this understanding, it is still worthwhile to observe the agreement among these statistical judges. For our purposes, we chose two essays to focus upon, written about one month apart from each other. One was written on the question of whether the "best things in life were really free," and the other on the "uses of anger." These will be called Essay C ("Free") and Essay D ("Anger"). For Essay C, the interjudge agreement is shown in the upper-left quadrant of Table IV-1.

Here the kind of agreement among judges is shown which is usually found for independent, subjective evaluations where there has been a certain amount of coaching, here ranging around .50 correlation of each individual judge with his peer.

It is to be expected that increasing the number of judges will increase the correlations, since it eliminates some random error from the judgments. To demonstrate this improvement, we have combined the columns of judgments in various ways, to find the effect of increasing the judges to two. When Column 1 (standing for the first columns of ratings for Essay C) is pooled with Column 2, this sum, shown in Column 5, may be correlated that of C3 + C4, shown in Column 6. The discovered correlation is .66, clearly higher than that between any two columns considered singly.

Additional comparisons may be made in a similar fashion, when the sum of C1 + C3 is correlated with the sum for the other columns. In fact, it is obvious that $\binom{4}{2} = 6$ such comparisons can be made, and the results (in natural order) are: .66, .67, .70, .70, .67, .66. A more complete listing of such intercorrelations, both between human judges, between human judge pairs, and between the single and combined columns, is shown in other cells of Table IV-1. Other parts of Table IV-1 will be discussed later in the chapter.

From psychometric theory, as well as from such empirical evidence, we would expect that the reliability of all four columns summed together would be higher still. When such a summation is done, however, it may no longer be correlated with others in the same fashion, since all of the data have been used. It is nevertheless possible to estimate such reliability through an analysis of variance

TABLE IV-1

HUMAN JUDGE AND JUDGE PAIR
CORRELATIONS FOR ESSAY C QUALITY
WITH PREDICTIONS FROM ESSAY D

	Variable	1	2	3	4	5	6	7	8	9	10	11	12	13
1.	C1		55	59	44	88	89	85	65	58	60	81	56	41
2.	C2	55		54	43	88	61	58	87	85	56	79	59	46
3.	C3	59	54		50	64	89	64	88	61	87	83	60	47
4.	C4	44	43	50		49	52	84	53	84	86	74	42	33
5.	C1 + C2	88	88	64	49		85	81	87	82	66	91	65	49
6.	C1 + C3	89	61	89	52	85		84	86	67	82	92	65	49
7.	C1 + C4	85	58	64	84	81	84		70	84	86	92	58	44
8.	C2 + C3	65	87	88	53	87	86	70		83	82	93	68	53
9.	C2 + C4	58	85	61	84	82	67	84	83		84	91	60	47
10.	C3 + C4	60	56	87	86	66	82	86	82	84		91	59	47
11.	All C	81	79	83	74	91	92	92	93	91	91		68	53
12.	All D	56	59	60	42	65	65	58	68	60	59	68		60
13.	D-Pred.	41	46	47	33	49	49	44	53	47	47	53	60	

of the columns, and such analysis was reported by McColly and Remstad, producing a reliability coefficient of about .83 for each of the two essays we are concerned with. Such a reliability is not very impressive for such an expensive rating process, but it is typical of such evaluation, and it does furnish an adequate target for the multiple regression of the proxes.

Having two different essays from each student writer, we may collect a certain amount of information about both individual and group stability across trials. Table IV-2 shows the means and standard deviations for the two student essays, first for Essay C ("Free"), then for Essay D ("Anger"). As explained previously, these proxes as shown here are not the raw frequencies for the essays, since such frequencies would have usually a large contaminating factor of essay length. Rather, they are the scores as converted to ratios and then multiplied to make a positive integer in each case. The transformation formulae are given in the FORTRAN program, printed in Appendix A.

The proxes employed have been previously described in Chapter III, and the reasoning employed for each, together with a prediction of the anticipated direction of correlation. These proxes were measured in the D essays, using an earlier version of the program listed here in Appendix A, and these proxes were then used in a multiple-regression analysis to predict the human judgments for Essay D. Among the aspects examined were the correlation of each prox with the criterion, the beta weight contributed by each prox, and the test-retest reliability of each prox. This information is summarized in Table IV-3.

In this table, Column A lists the proxes by title, and in the same order as described in the last chapter. Column B shows the correlation of each prox with the criterion, which was the sum of four human ratings for each essay. And Column D indicates the test-retest reliability

TABLE IV-2

MEANS AND STANDARD DEVIATIONS OF
THE PROX SCORES

	Proxes	Essay C		Essay D	
		Mean	St. Dev.	Mean	St. Dev.
1.	Title present	.90	.29	.83	.38
2.	Av. sentence length	176.79	41.91	175.48	41.85
3.	Number of paragraphs	5.47	2.14	5.16	1.90
4.	Subject-verb openings	49.27	14.01	45.01	12.36
5.	Length of essay in words	397.40	112.62	361.32	104.44
6.	Number of parentheses	2.02	4.55	1.67	4.09
7.	Number of apostrophes	9.62	9.20	8.75	8.28
8.	Number of commas	48.81	22.26	40.97	21.64
9.	Number of periods	56.70	12.36	58.25	12.20
10.	Number of underlined words	1.74	3.65	1.42	3.26
11.	Number of dashes	1.97	4.50	1.37	3.39
12.	No. colons	.57	1.59	.47	1.72
13.	No. semicolons	1.54	2.97	1.22	2.73
14.	No. quotation marks	293.19	275.16	114.45	156.48
15.	No. exclamation marks	8.61	23.36	11.06	34.00
16.	No. question marks	53.35	70.97	25.46	49.37
17.	No. prepositions	9.73	1.76	8.90	1.80
18.	No. connective words	.35	.51	.43	.58
19.	No. spelling errors	.11	.31	.12	.34
20.	No. relative pronouns	2.03	1.05	1.93	.96
21.	No. subordinating conjs.	2.22	1.00	2.89	1.14
22.	No. common words on Dale	81.89	4.58	79.17	5.09
23.	No. sents. end punc. pres.	99.07	3.23	99.52	1.82
24.	No. declar. sents. type A	92.45	8.48	95.48	6.35
25.	No. declar. sents. type B	.56	1.63	.61	2.11
26.	No. hyphens	2.53	4.69	1.95	3.94
27.	No. slashes	.05	.39	.10	.59
28.	Aver. word length in ltrs.	423.77	23.41	438.36	24.47
29.	Stan. dev. of word length	217.72	20.31	232.32	23.13
30.	Stan. dev. of sent. length	82.56	29.63	78.07	31.83

NOTE: These means and standard deviations are based upon the transformed scores, altered so that every individual score would be a positive integer, and would usually express a relative rather than an absolute frequency.

TABLE IV-3

PROXES USED TO PREDICT A CRITERION OF OVERALL QUALITY (ESSAY D)

	A. Proxes	B. Corr. with Criterion	C. Beta Wts.	D. Test-Ret. Rel. (Two essays)
1.	Title present	.04	.09	.05
2.	Av. sentence length	.04	-.13	.63
3.	Number of paragraphs	.06	-.11	.42
4.	Subject-verb openings	-.16	-.01	.20
5.	Length of essay in words	.32	.32	.55
6.	Number of parentheses	.04	-.01	.21
7.	Number of apostrophes	-.23	-.06	.42
8.	Number of commas	.34	.09	.61
9.	Number of periods	-.05	-.05	.57
10.	Number of underlined words	.01	.00	.22
11.	Number of dashes	.22	.10	.44
12.	No. colons	.02	-.03	.29
13.	No. semicolons	.08	.06	.32
14.	No. quotation marks	.11	.04	.27
15.	No. exclamation marks	-.05	.09	.20
16.	No. question marks	-.14	.01	.29
17.	No. prepositions	.25	.10	.27
18.	No. connective words	.18	-.02	.24
19.	No. spelling errors	-.21	-.13	.23
20.	No. relative pronouns	.11	.11	.17
21.	No. subordinating conjs.	-.12	.06	.18
22.	No. common words on Dale	-.48	-.07	.65
23.	No. sents. end punc. pres.	-.01	-.08	.14
24.	No. declar. sents. type A	.12	.14	.34
25.	No. declar. sents. type B	.02	.02	.09
26.	No. hyphens	.18	.07	.20
27.	No. slashes	-.07	-.02	-.02
28.	Aver. word length in ltrs.	.51	.12	.62
29.	Stan. dev. of word length	.53	.30	.61
30.	Stan. dev. of sent. length	-.07	.03	.48

*Number of students judged was 272. Multiple R against human criterion (four judges) was .71 for both Essay C and Essay D (D data shown here). F-ratios for Multiple R were highly significant.

for the proxes, that is, the correlation between Essay C and Essay D for the proxes, as a measure of writing habit, or stability of writing behavior, in the student writers.

Overall prediction of the proxes. In multivariate analysis, it is often pointless to elaborate a hypothesis for each predictor, and to explain how each variable met expectations, or failed to do so. But it may be instructive to note how well the predictions fared as a whole. While some of the predictions were very tentative and loose, and while many of the variables obviously had only a non-significant relation with the criterion, some estimate may be made of the overall success of the predictions.

In general, the predictions were quite accurate, notwithstanding the obvious large random errors in the relationships which are evident in the table. The degree of success was examined and the results are shown in Table IV-4, which displays a contingency diagram for the direction of prox correlation with the criterion (positive or negative direction), and shows the relation between the predicted and discovered directions. Here the number of agreements is seen as 21, and disagreements 7. As is also shown in Table IV-4, the chi square was computed to be 3.12, which, with one degree of freedom and the assumption that a one-tailed test is appropriate for such agreement, is significant at the five per cent level of confidence. One may conclude, therefore, that most the predictions were significantly in the correct direction.

Correlation with the criterion. It does not make much sense to describe a summary table in any detail, but it is useful to comment on a few outstanding points. As was explained in the last chapter, many of the predictors used were "proxes of opportunity", and it is not surprising that they were relatively unproductive. This is generally true for the large number of punctuation marks. The more major contributors to empirical prediction were usually foreseen.

TABLE IV-4

THE DIRECTION OF CORRELATIONS OF PROXES
WITH THE CRITERION:
PREDICTED AND OBSERVED FREQUENCIES

		Predicted	
		+	-
Observed	+	17	1
	-	6	4

N = 28, since two variables were not predicted.

$$X^2 = \frac{N(AD - BC - \frac{N}{2})^2}{(A+B)(C+D)(A+C)(B+D)} = 3.12 \text{ (significant).}$$

TABLE IV-5

INTERCORRELATIONS OF THE PROXES FOR ESSAY D

Prox	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1		-09	04	01	02	-23	-05	-07	01	02	03	05	-01	01	04	03	-01	05	07	-04	03	10	02	-05	01	06	02	-12	-12	-07
2	-09		-23	-12	07	03	-02	10	-88	-06	-06	-01	11	07	-00	-04	21	-04	-02	21	12	-04	11	06	-01	-08	01	01	02	70
3	04	-23		08	36	12	-03	-09	19	10	19	-02	-04	07	05	03	05	00	-08	-04	-11	-01	-22	-11	-05	-02	04	-01	06	-10
4	01	-12	08		-06	-07	10	-20	10	14	-08	-04	-11	-19	07	-06	-16	-05	-04	13	-14	21	01	05	-08	-04	-17	-18	-20	-08
5	02	07	36	-06		05	00	06	-08	00	09	-03	-08	02	-00	-07	04	-02	-14	03	00	-03	14	09	01	02	04	-02	03	02
6	-23	03	12	-07	05		16	15	-05	18	15	17	14	23	23	11	03	02	06	02	-09	-06	-18	-21	00	10	06	02	09	21
7	-05	-02	-03	10	00	16		-03	-02	13	04	10	03	05	22	13	-27	-09	-02	02	06	36	01	-20	01	-08	03	-30	-29	12
8	-07	10	-09	-20	06	15	-03		-15	09	24	14	14	26	06	08	13	15	-18	09	-15	-30	00	-14	12	19	-06	30	33	03
9	01	-88	19	10	-08	-05	-02	-15		-00	02	-01	-07	-14	-13	-15	-23	04	-00	-21	-07	01	-00	17	-00	09	-01	02	00	-59
10	02	-06	10	14	00	18	13	09	-00		10	05	-06	11	16	01	04	01	-02	10	-16	-01	-02	-11	-03	05	-03	-03	-04	03
11	03	-06	19	-08	09	15	04	24	02	10		20	01	19	04	07	-00	13	-11	-02	-10	-17	-17	-15	10	10	-06	15	14	03
12	05	-01	-02	-04	-03	17	10	14	-01	05	20		15	21	10	02	06	08	-05	00	-01	-04	-12	-10	-00	-02	01	-02	-03	06
13	-01	11	-04	-11	-08	14	03	14	-07	-06	01	15		13	18	01	-03	05	-06	-01	-06	-10	-00	-05	02	-07	-07	09	05	14
14	01	07	07	-19	02	23	05	26	-14	11	19	21	13		16	12	-05	-00	04	10	05	-08	-13	-25	16	04	04	09	06	14
15	04	-00	05	07	-00	23	22	06	-13	16	04	10	18	16		12	-08	-06	00	-00	-10	14	-06	-47	-00	10	-05	-18	-21	09
16	03	-04	03	-06	-07	11	13	08	-15	01	07	02	01	12	12		-09	03	01	-05	-12	06	-03	-80	-05	09	-08	-11	-10	30
17	-01	21	05	-16	04	03	-27	13	-23	04	-00	06	-03	-05	-08	-09		17	-05	-00	-27	-35	-06	06	04	01	01	26	28	04
18	05	-04	00	-05	-02	02	-09	15	04	01	13	08	05	-00	-06	03	17		-12	06	-14	-25	-06	-00	-06	01	-02	26	28	-05
19	07	-02	-08	-04	-14	06	-02	-18	-00	-02	-11	-05	-06	04	00	01	-05	-12		-09	06	-02	-08	-05	01	-04	18	02	-03	-05
20	-04	21	-04	13	03	02	02	09	-21	10	-02	00	-01	10	-00	-05	-00	06	-09		-13	08	-01	03	-00	-10	-05	-01	-01	12
21	03	12	-11	-14	00	-09	06	-15	-07	-16	-10	-01	-06	05	-10	-12	-27	-14	06	-13		29	01	17	-07	-04	08	-23	-25	-03
22	10	-04	-01	21	-03	-06	36	-30	01	-01	-17	-04	-10	-08	14	06	-35	-25	-02	08	29		02	-07	-04	-17	09	-82	-77	03
23	02	11	-22	01	14	-18	01	00	-00	-02	-17	-12	-00	-13	-06	-03	-06	-06	-08	-01	01	02		31	04	06	02	-03	-06	01
24	-05	06	-11	05	09	-21	-20	-14	17	-11	-15	-10	-05	-25	-47	-80	06	-00	-05	03	17	-07	31		-27	-09	09	14	13	-27
25	01	-01	-05	-08	01	00	01	12	-00	-03	10	-00	02	16	-00	-05	04	-06	01	-00	-07	-04	04	-27		-05	-01	02	-01	00
26	06	-08	-02	-04	02	10	-08	19	09	05	10	-02	-07	04	10	09	01	01	-04	-10	-04	-17	06	-09	-05		-05	15	16	-05
27	02	01	04	-17	04	06	03	-06	-01	-03	-06	01	-07	04	-05	-08	01	-02	18	-05	08	09	02	09	-01	-05		-06	-03	-01
28	-12	01	-01	-18	-02	02	-30	30	02	-03	15	-02	09	09	-18	-11	26	26	02	-01	-23	-82	-03	14	02	15	-06		86	-11
29	-12	02	06	-20	03	09	-29	33	00	-04	14	-03	05	06	-21	-10	28	28	-03	-01	-25	-77	-06	13	-01	16	-03	86		-11
30	-07	70	-10	-08	02	21	12	03	-59	03	03	06	14	14	09	30	04	-05	-05	12	-03	03	01	-27	00	-05	-01	-11	-11	

TABLE IV-6

INTERCORRELATIONS OF THE PROXES FOR ESSAY C

Prox	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1		-11	-02	-04	-11	01	-05	08	08	02	-05	04	-14	-04	-01	06	-01	-00	03	-18	-06	-03	03	-03	02	-07	04	02	-00	-09
2	-11		-14	07	13	01	-03	03	-85	-05	04	-00	10	30	-07	-19	10	-01	06	27	21	-06	-03	17	-12	-01	-02	01	02	67
3	-02	-14		-04	47	05	01	06	13	05	17	-06	-01	03	04	-04	11	-00	02	-07	-01	-03	-29	-11	08	-05	-05	-00	04	-05
4	-04	07	-04		-01	-05	-03	-23	-02	-05	-04	-00	-02	-10	-07	-21	-04	-28	04	15	-02	13	03	20	-03	-10	01	-05	-10	-06
5	-11	13	47	-01		-01	-02	05	-09	-00	16	-04	05	08	04	-07	-01	-00	-06	06	-00	03	01	04	-04	-02	-00	-08	03	14
6	01	01	05	-05	-01		08	08	-02	08	07	12	06	13	-02	-05	-03	-01	-00	-02	07	-07	-29	-04	-07	04	-04	03	08	07
7	-05	-03	01	-03	-02	08		16	04	17	-01	10	01	-07	18	08	-11	-14	05	-01	05	20	-04	-12	03	12	08	-21	-14	08
8	08	03	06	-23	05	08	16		-09	15	16	-11	10	18	18	12	19	08	04	-14	-13	-30	-04	-14	-05	34	04	23	27	18
9	08	-85	13	-02	-09	-02	04	-09		-04	-12	-05	-13	-32	-01	-13	-11	04	-05	-22	-21	05	15	15	12	-03	-01	01	-01	-59
10	02	-05	05	-05	-00	08	17	15	-04		03	19	-03	03	16	19	-01	-02	15	01	-12	05	-09	-25	-02	07	31	-08	-04	-00
11	-05	04	17	-04	16	07	-01	16	-12	03		08	09	17	00	18	03	-02	-09	02	-03	-06	-02	-17	10	13	-05	08	13	07
12	04	-00	-06	-00	-04	12	10	11	-05	19	08		08	09	12	12	07	04	-04	-03	-04	-03	02	-13	-05	13	13	-00	-01	11
13	-14	10	-01	-02	05	06	01	10	-13	-03	09	08		13	08	06	10	07	-01	-03	-08	-12	05	-06	05	-03	-03	12	15	13
14	-04	30	03	-10	08	13	-07	18	-32	03	17	09	13		00	04	13	05	-02	02	01	-10	-03	-05	04	17	-02	04	07	26
15	-01	-07	04	-07	04	-02	18	18	-01	16	00	12	08	00		15	07	-00	13	-16	-13	-00	03	-37	-01	06	01	-07	-06	07
16	06	-19	-04	-21	-07	-05	08	12	-13	19	18	12	06	04	15		-00	-01	08	-09	-07	04	01	-86	05	07	10	-09	-07	04
17	-01	10	11	-04	-01	-03	-11	19	-11	-01	03	07	10	13	07	-00		-06	11	-19	-23	-31	00	-00	-07	16	07	19	21	11
18	-00	-01	-00	-28	-00	-01	-14	08	04	-02	-02	04	07	05	-00	-01	-06		05	04	02	-08	11	06	-05	-04	-01	11	07	-05
19	03	06	02	04	-06	-00	05	04	-05	15	-09	-04	-01	-02	13	08	11	05		-02	-05	03	04	-07	-02	02	02	-07	-10	00
20	-18	27	-07	15	06	-02	-01	-14	-22	01	02	-03	-03	02	-16	-09	-19	04	-02		06	10	05	14	-06	-01	12	-04	-11	12
21	-06	21	-01	-02	-00	07	05	-13	-21	-12	-03	-04	-08	01	-13	-07	-23	02	-05	06		15	-12	05	01	-11	00	-09	-13	12
22	-03	-06	-03	13	03	-07	20	-30	05	05	-06	-03	-12	-10	-00	04	-31	-08	03	10	15		-02	-05	01	-31	01	-81	-76	-13
23	03	-03	-29	03	01	-29	-04	-04	15	-09	-02	02	05	-03	03	01	00	11	04	05	-12	-02		37	-08	03	04	03	04	-03
24	-03	17	-11	20	04	-04	-12	-14	15	-25	-17	-13	-06	-05	-37	-86	-00	06	-07	14	05	-05	37		-25	-05	-07	11	09	-05
25	02	-12	08	-03	-04	-07	03	-05	12	-02	10	-05	05	04	-01	05	-07	-05	-02	-06	01	01	-08	-25		-07	-02	-01	02	-10
26	-07	-01	-05	-10	-02	04	12	34	-03	07	13	13	-03	17	06	07	16	-04	02	-01	-11	-31	03	-05	-07		-02	26	19	14
27	04	-02	-05	01	-00	-04	08	04	-01	31	-05	13	-03	-02	01	10	07	-01	02	12	00	01	04	-07	-02	-02		01	04	-01
28	02	01	-00	-05	-08	03	-21	23	01	-08	08	-00	12	04	-07	-09	19	11	-07	-04	-09	-81	03	11	-01	26	01		82	08
29	-00	02	04	-10	03	08	-14	27	-01	-04	13	-01	15	07	-06	-07	21	07	-10	-11	-13	-76	04	09	02	19	04	82		09
30	-09	67	-05	-06	14	07	08	18	-59	-00	07	11	13	26	07	04	11	-05	00	12	12	-13	-03	-05	-10	14	-01	08	09	

The average sentence length was anticipated, from other literature, to be more important in a multivariate than bivariate way, and this is apparently the case. But these higher-order relationships, which are only hinted by the beta weights and intercorrelations, are very difficult to articulate. The length of essay in words, number of commas, number of prepositions, number of connectives, relative pronouns, spelling errors, common words, and long words were all in the anticipated direction.

On the other hand, there were some surprises in the data. The number of question marks was predicted to be indicative of variety in style, yet was a negative predictor. On the second set of essays analyzed, however, it has moved from -.14 to .08, which implies that there may be an interaction of this feature with the wording of the assigned topics, or of the accompanying instructions to the student writers. Such an interaction becomes plausible in light of the interrogative wording of the "Free" question.

Another surprise was the negative correlation with the criterion fo variable 21, the proportion of subordinating conjunctions. The assumption that the proportion would reflect complexity, and that complexity would be related to maturity of style, was not destroyed, but it surely was shadowed by the negative correlation of -.12 for the D essays. Here an interaction with topic is not a plausible explanation, since for Essay C the discovered correlation had moved only slightly, to -.06. It _is_ worth note that for both essays, when the _other_ predictors are taken into consideration, the beta weights for subordinating conjunctions are both positive. But here again, explanation of such higher-order effects are difficult to ascertain. It is probable that the explanations of this surprise should be pursued in the specific words in the list of subordinating conjunctions, and in further syntactic analysis of the sentences where they are used. This sort of exploration is further discussed in the final chapter of this report.

Surely the question of fluency is a most important one in the evaluation of essays. Two strong arguments for some general trait of prolixity appear in the importance of word length and essay length -- the first being the highest correlate with the criterion, the second yielding the highest beta weight. And these relative positions are maintained for the C essays, as well. So that such comparisons may be easily made, Table IV-7 contains the prox information for Essay C, the "free" essay. Column A has of course the title of the proxes. Column B shows again the correlation with the criterion. And Column C displays the beta weights for the proxes, when all variables are used to maximize the prediction of overall human judgment.

This table (IV-7) has the same status as Table IV-3, for the D essays. The D essays were presented first only because, historically, they were analyzed first. In fact, what they have in common is at once apparent to the naked eye. Most of the important correlations with the criterion are maintained in Table IV-7, and most of the important beta weights have sustained their contributions with the second essays.

Multiple regression. From the standpoint of overall simulation, the multiple correlation obtainable for the pooled human judgments is the primary goal of the analysis. For Essay D, the multiple-R achieved was a rather startling .71. And when it was possible to perform the same analysis for Essay C, although there were obvious changes as we have seen, the resultant multiple-R was once more (coincidentally) just .71. This coefficient means that for this set of proxes, and for these sets of essays, the correlation between the human ratings actually achieved, and the "predicted" ratings generated by the discovered beta vector, would be .71. Given the looseness of human rating, and the pooled human reliability of only .83, the multiple regression coefficient is encouraging in the extreme.

TABLE IV-7

PROXES USED TO PREDICT A CRITERION

OF OVERALL QUALITY (ESSAY C)

	A. Proxes	B. Corr. with Criterion	C. Beta wts.
1.	Title present	.03	.06
2.	Av. sentence length	-.07	.09
3.	Number of paragraphs	.08	-.02
4.	Subject-verb openings	-.01	.09
5.	Length of essay in words	.25	.03
6.	Number of parentheses	-.05	-.05
7.	Number of apostrophes	-.16	-.09
8.	Number of commas	.36	-.29
9.	Number of periods	.01	-.01
10.	Number of underlined words	-.06	.07
11.	Number of dashes	.31	-.15
12.	No. colons	.14	-.06
13.	No. semicolons	.09	.17
14.	No. quotation marks	.12	-.12
15.	No. exclamation marks	-.04	-.09
16.	No question marks	.08	-.05
17.	No. prepositions	.16	-.06
18.	No. connective words	.11	.10
19.	No. spelling errors	-.21	.01
20.	No. relative pronouns	.01	.10
21.	No. subordinating conjs.	-.06	.25
22.	No. common words on Dale	-.37	.15
23.	No. sents. end punc. pres.	.12	.34
24.	No. declar. sents. type A	-.00	-.05
25.	No. declar. sents. type B	-.05	.11
26.	No. hyphens	.26	-.10
27.	No. slashes	.03	.00
28.	Aver. word length in ltrs.	.37	-.03
29.	Stan. dev. of word length	.45	.26
30.	Stan. dev. of sent. length	.08	.09

As is well known, however, we should not expect all of this accuracy if we took new essays and applied the discovered beta weightings to them, to predict their human ratings. For any set of scores, or any set of resultant correlations, contains not only true variance associated with the variable, but also a certain amount of error variance, random for the particular subjects concerned, which will not ordinarily be found with a new set of human subjects, or essays. The true variance gives us information which will be subsequently useful. But the error variance is also capitalized upon by the analysis, and a certain portion of the multiple-regression coefficient, and of the contributing beta weights, will spuriously seem to contribute, but will not stand up in a replication.

When one does run such an analysis, then, and subsequently cross-validates the weightings with new data, the resulting predictions will not correlate as highly with the criterion as one might hope. The statistical loss is commonly spoken of as "shrinkage" and has been widely treated in the literature (e.g., McNemar, 1962). Fortunately, empirical cross-validation is not always necessary, since the performance of such data may partly be predicted mathematically. As one would suppose, the larger the number of subjects, the more reliable the multiple-R will be; but the larger the number of variables (given the same number of subjects), the <u>less</u> reliable the multiple-R will be.

<u>The Paulus tables</u>. Since our work of essay analysis continues to be heavily dependent upon multiple regression, Dieter Paulus has made an investigation of the behavior of such data, given a varying N of subjects, and varying <u>n</u> of variables. Some of his findings are set forth in a usable form in Tables IV-8 and IV-9. Table IV-8 shows the minimum Multiple R coefficients required for significance

TABLE IV - 8

MINIMUM MULTIPLE CORRELATIONS REQUIRED
FOR SIGNIFICANCE AT THE .05 LEVEL

SAMPLE SIZE

NUMBER PREDICTORS	50	75	100	125	150	175	200	250	300	400
5	.4652	.3815	.3308	.2963	.2703	.2509	.2346	.2099	.1916	.1659
10	.5898	.4861	.4221	.3788	.3460	.3215	.3008	.2694	.2459	.2131
15	.6828	.5635	.4901	.4426	.4047	.3746	.3498	.3143	.2870	.2495
20	.7565	.6282	.5485	.4942	.4513	.4180	.3925	.3521	.3217	.2790
25	.8207	.6858	.5994	.5388	.4927	.4578	.4290	.3851	.3520	.3046
30	.8751	.7337	.6429	.5778	.5301	.4929	.4622	.4151	.3796	.3287
35	.9227	.7790	.6831	.6166	.5641	.5249	.4924	.4415	.4039	.3498
40	.9623	.8189	.7203	.6492	.5958	.5536	.5183	.4648	.4265	.3707
45	.9923	.8568	.7549	.6812	.6261	.5809	.5455	.4897	.4471	.3875
50		.8906	.7875	.7118	.6540	.6074	.5694	.5115	.4684	.4063

TABLE IV - 9

MINIMUM MULTIPLE CORRELATIONS REQUIRED
FOR SIGNIFICANCE AT THE .01 LEVEL

SAMPLE SIZE

NUMBER PREDICTORS	50	75	100	125	150	175	200	250	300	400
5	.5312	.4388	.3819	.3433	.3135	.2907	.2724	.2444	.2231	.1933
10	.6471	.5382	.4705	.4234	.3871	.3599	.3369	.3021	.2764	.2396
15	.7322	.6107	.5345	.4837	.4437	.4114	.3839	.3452	.3160	.2741
20	.7996	.6726	.5901	.5327	.4893	.4532	.4256	.3831	.3502	.3033
25	.8572	.7257	.6397	.5764	.5293	.4917	.4621	.4155	.3809	.3301
30	.9042	.7703	.6802	.6134	.5640	.5254	.4931	.4447	.4070	.3520
35	.9444	.8116	.7184	.6510	.5966	.5574	.5247	.4728	.4319	.3739
40	.9762	.8488	.7529	.6824	.6286	.5847	.5484	.4933	.4528	.3931
45	.9968	.8825	.7852	.7125	.6575	.6113	.5738	.5166	.4734	.4122
50		.9129	.8151	.7408	.6829	.6355	.5958	.5390	.4943	.4296

(at the 5% confidence level) with different n and N. The number of predictors is scaled from 5 to 50, along the left hand column, and the number of subjects is scaled from 50 to 400, along the top. It may be easily seen, then, that for the present investigation, where predictors number 30 and cases just over 250, a multiple-R of about .41 is necessary for significance at the .05 level.

Table IV-9 shows similar requirements for the .01 level of confidence, showing that around .44 is necessary to reject the null hypothesis. These Paulus tables are very convenient in dealing with large numbers of such coefficients, and seem to be a useful by-product of the present research. (For the computational reasoning, see Kelley, 1947, p. 475.)

There is another familiar problem in interpreting regression, however, and this one depends on the reliability of the criterion. It is obviously impossible to predict perfectly a criterion which is itself not perfectly reliable. And the reliability of a group of human raters obviously depends on the number of such raters and on their inter-judge agreement. As we have seen, the reliability of the group of four raters in Wisconsin was .83, and this means that about 31% of the variance ($1.00 - .83^2$) would be unexplained and indeed "unpredictable." When one is considering purely practical predictions for groups that are identical, it is reasonable to ignore this handicap. But when one is attempting to assess the "true" accuracy of a set of predictors, it is more fair to take such criterion unreliability into consideration.

Paulus designed Table IV-10 to do just this task. The left column refers to the discovered multiple-R coefficient, and the top headings refer to the measured reliability of the criterion variable. Just by finding the appropriate cell of this table, then, one may infer what

TABLE IV-10

MULTIPLE REGRESSION COEFFICIENTS

CORRECTED FOR ATTENUATION

CAUSED BY CRITERION UNRELIABILITY

Discovered MULTR Coefficient	Reliability of Criterion Variable .35	.40	.45	.50	.55	.60	.65	.70	.75	.80
.50	85	79	75	71	67	65	62	60	58	56
.51	86	81	76	72	69	66	63	61	59	57
.52	88	82	78	74	70	67	65	62	60	58
.53	90	84	79	75	71	68	66	63	61	59
.54	91	85	81	76	73	70	67	65	62	60
.55	93	87	82	78	74	71	68	66	64	61
.56	95	89	83	79	76	72	69	67	65	63
.57	96	90	85	81	77	74	71	68	66	64
.58	98	92	86	82	78	75	72	69	67	65
.59	99	93	88	83	80	76	73	71	68	66
.60		95	89	85	81	77	74	71	69	67
.61		96	91	86	82	79	76	73	70	68
.62		98	92	88	84	80	77	74	72	69
.63		99	94	89	85	81	78	75	73	70
.64			95	91	86	83	79	76	74	72
.65			97	92	88	84	81	78	75	73
.66			98	93	89	85	82	79	76	74
.67			99	95	90	87	83	80	77	75
.68				96	92	88	84	81	79	76
.69				98	93	89	86	82	80	77
.70				99	94	90	87	84	81	78
.71					96	92	88	85	82	79
.72					97	93	89	86	83	81
.73					98	94	91	87	84	82
.74					99	96	92	88	85	83
.75						97	93	90	87	84
.76						98	94	91	88	85
.77						99	96	92	89	86
.78							97	93	90	87
.79							98	94	91	88
.80							99	96	92	89

the discovered correlation <u>might</u> have been if the criterion had been perfectly reliable. This table was produced, like the two before, from equations programmed by Paulus for the time-sharing console in the Bureau of Educational Research at Connecticut. It was based upon the division of the multiple-R coefficient by the square root of the reliability of the criterion variable (Kelley, 1947, p. 412).

Still another table serving such needs was designed to perform automatic "shrinkage" of multiple-R coefficients. As we have noted, when MULTR is calculated, it finds a maximum fit of weightings to the sample data. But the sample data do not reflect merely the true covariances of the population. They also reflect random error typical only of the cases constituting the sample, and the computational method capitalizes upon such random error, just as it capitalizes upon the true covariance. And such random error increases rapidly as $\underline{n}$, the number of predictor variances, increases. As we have also noted, however, the sample size tends to counteract this mounting random error. The "shrunken" multiple-regression coefficient, then, is the statistical estimate of what the coefficient would have been, if it had not capitalized on such random error. It is therefore, of course, always smaller than the observed coefficient.

There are several formulas available for such shrinkage. Perhaps the most appropriate one is the Wherry formula (Kelley, 1947, p. 474), expressed by:

$$R_s^2 = \frac{(N - 1)\ R^2 - n}{N - n - 1}$$

where R_s is defined as the shrunken coefficient, R is the discovered coefficient in the sample, N is the number of persons cases in the sample, and n is the number of predictor variables. The Wherry formula expresses what is believed to be the "true" multiple-R coefficient in the population of interest (rather than in some other sample from that population). This formula thus seems most appropriate for such exploratory research, where the population parameters are indeed of central interest. And it was therefore programmed into the computer to produce the various tables for shrinkage.

These tables are listed in Appendix B, since they are too large to include conveniently in the running text. In Appendix B, the tables are divided according to size of n, the number of predictor variables. These sub-tables are as follows:

Organization of Table IV-11
(See Appendix B)

Sub-Table	Number of Predictors
A	25
B	30
C	35
D	40
E	45
F	50
G	55

To avoid too massive a document, Paulus restricted the size of **n**, therefore, to the range from 25 to 55. He also restricted the sample size to a range from 100 to 300. Both of these constraints mean that the tables are very appropriate for studies of the present size, and a large number of empirical studies seem to fall within these limits.

Use of the tables. In the present case, we can immediately apply certain of these tables to the discovered data. We would ordinarily shrink the MULTR before we would correct it for attenuation; therefore we would enter Table IV-11(B), with n = 30, N greater than 275, and a discovered R of .71. The appropriate cell yields us a shrunken R of .67. Then we may enter Table IV-10 with a new "discovered R" of .67, and a known criterion reliability of over .80. In the appropriate cell we find a coefficient of .75, when first shrunken and then corrected for attenuation.

In this use as in other uses of such tables, the user should always remember that these table cells are generated from formulae which inevitably make certain theoretical assumptions about the distribution of the data. One will not necessarily expect, for instance, that cross-validation with another sample will match the Wherry shrinkage with any exactness. In the first place, violations of assumed distributions will often cause a greater shrinkage through cross-validation than one would expect. On the other hand, the prediction of a new sample will often be, for the reasons touched on above, lower than the corresponding prediction would be of the population itself. But these tables can surely supply rather good approximations to the statistics which we may be very much interested in, but cannot measure directly.

Reliability of proxes. To some extent, validation with subsequent samples will depend upon the reliability of the multiple correlation, and this will depend in part upon the reliability of the individual proxes. As already noted, the reliabilities of these proxes are shown in Column D of Table IV-3. The coefficients of Column D are the product-moment correlations between the different essays for a particular prox. For example, for the second prox listed, "average sentence length," the coefficient represents the similarity between these averages for two different essays written about one month apart. Thus the correlation is an extremely conservative one, and seems a reasonable measure of "writing behavior" under two separate (but quite similar) stimulus situations.

A few generalizations may be made about the prox reliabilities. In the first place, it seems that there is a correlation between Column B and Column D. Those proxes with the highest reliability are also typically those which aid most in the prediction of overall quality. The highest reliabilities seem to belong (in descending order of magnitude) to: the proportion of common words (#22), average sentence length (#2), average word length (#28), proportion of commas (#8), and length of essay in words (#5). With the exception of average sentence length, these same proxes are among the best (bivariate) predictors of writing quality, and even average sentence length is among the more substantial contributors in the combined, multivariate prediction.

A second generalization about such proxes is that their reliabilities may be related to the frequency of occurrence. Those proxes which deal with the most frequent events, such as average length of word, or proportion of common words, may have the highest reliability. Sentences, which are also found in a fair number within an essay, have a fairly stable reliability for average length (.63). And paragraphs, which are less frequent in an essay than sentences, have a frequency reliability which is somewhat lower (.42). On the other hand, the writing of a title, which is a behavioral decision which occurs only once in the writing of each essay, has a practically non-existent reliability. This is a generalization which is still very tentative, and deserving of more exploration.

A third generalization, really a speculation, is that there may be a significant interrelationship among the reliability of the prox, the beta weight of the prox for a particular essay, and the worthiness of the prox for assessing the more stable writing behavior of the student. It is worth studying to find out whether prediction of future essays may be improved by modifying the beta weights in accordance with the reliability of the proxes. This possibility has not yet been analyzed within this project, but is promising for the future, for practical prediction purposes.

A final interesting speculation concerns the relationship between the reliabilities of the proxes and the reliability of the total multiple-regression equation. It is a familiar observation in mental testing that the total score of a test, which is often a sum of various part scores, will frequently be more reliable than any of those part scores taken separately. But it might not be so obvious that the same phenomenon may occur in multiple regression, that the total predictive validity may conceivably be higher than the reliability of any of the contributing predictors. This appears to be the case here; but the mathematical aspects of this problem will not be analyzed within the scope of this report.

Human and machine judgments. Now it would be valuable to return to a further analysis of Table IV-1, since it has much to tell us about rater performance. Earlier in this chapter, it was noted that the upper left quadrant (for Columns 1-4) shows us the intercorrelations among the judge columns for Essay C. Columns 5 through 10 show the increased accuracy, or reliability, which may come from increasing the number of judges.

In this portion of the table, many of the coefficients are of course inflated artificially through a part-whole agreement. Columns 1 and 5, for example, agree at a level of .88, but since Column 5 is simply the sum of Column 1 and Column 2, this has little empirical meaning.

whether a coefficient is so contaminated may be at once determined by reading the variable names in the leftmost column of the table.

Column 11 represents the sum of all C ratings, and therefore the agreement coefficients between 11 and all of the earlier columns are similarly a part-whole artifact. On the other hand, Columns 11, 12, and 13 do have considerable significance when properly understood. Column 12 represents the sum of all D ratings, and is therefore the best measure of external validity which one could wish for the various ratings given by the human judges to the C essays. And Column 13 represents the machine evaluation, derived from multiple-regression analysis of the proxes for Essay D. This information has particular meaning for this project, as is here explained.

Human vs. machine "validity". An interesting sidelight is cast upon the human vs. machine by looking at some analyses of the human judges of essay C compared with human and machine judgments of essay D. The most important meaning of "validity", for an essay test, would appear to be how well it predicts performance on another essay by the same student writers. That is, in the long run we are less interested in how reliable this particular judgment of performance is, and more interested in how well it assesses the student's general writing performance, under somewhat differing circumstances. One important measure of this validity, then, would be agreement of ratings with those of other essays by the same students.

We would always expect such validities to be lower than the agreement between raters on the same essays, for not only would the ratings differ because of rater error (or viewpoint), but they would differ also because of intrinsic differences in performance of the student under two sets of conditions. One interesting comparison of the machine and the human judge, then, would be to match each with the ratings of the expert group for some second essay.

This comparison was simplified for the present study because, as we have seen, multiple essays were analyzed for the same student writers, the "Free" essays and the "Anger" essays. We have seen how the individual judges (or their statistical summations) agreed with each other in Table IV-1, on the C essays. Now it would be instructive to see how well each of those "individual" judges predicted the ratings on the D essays. For this comparison they are correlated with those ratings given by the group of four raters on the "Free" essays, so that their coefficients each represent the correlation of an individual with a group of individuals. And we may therefore expect the correlations to be higher than those between pairs of individuals, since some of the error (but by no means all) will be eliminated by the larger number involved in the group sums. The coefficients are also somewhat higher then they should be, in one sense, since some of the same judges were involved in evaluating both C and D essays.

When such comparisons are made, the four judges of C are found to correlate with the pooled judgment of D as follows: .56, .59, .60, and .42. These coefficients produce an average of about .54 between these two essays.

On the other hand, we could look at the predictions of the D essays generated by prox analysis of these same essays, resulting from the multiple regression programs. These predictions become a (reasonably) independent way of estimating how well the student might do on another essay. When these machine predictions of D, then, are used to predict the students' actual performance on the C essays (that is, the pooled expert judgment of such performance), the coefficient is .53. This coefficient is almost precisely that of the typical human judge, and once again shows us how similar to the human individual is this first approximation of a machine system.

This finding also furnishes another response to the critic who supposes that the measures found through such statistical procedures are entirely artifactual, and will disappear upon validation. There could hardly be any measure of validity of essay rating superior to this one, and on this measure the machine performs, even in this early state, as well as the expert human.

Human vs. machine accuracy. There are two elements of "accuracy" of rating which are submerged in the data, and for which there is no readily available statistic in common use. Both of these hidden statistics are extremely important and, as it happens, both would argue additional advantages on the side of machine grading, in almost any practical situation.

When students speak of "fairness" in grading, they ordinarily are not speaking primarily of any correlation with some true score, as much as they are speaking of absolute comparisons with such a true score. As we know, the common correlational methods suppress both mean scores and score variances, in order to make the comparison on standard scores alone. Therefore it would be possible to have two human raters "agree" perfectly, in terms of correlation coefficient, in that *r* might equal 1.00. Yet they might have not one rating in common. This would be the case if two teachers assigned the identical rank orders to a set of students, yet one assigned marks just one grade lower than the other teacher. To the typical student, such a question of "hard" or "easy" marking would be much more important than minor differences in correlation.

Another aspect of accuracy or "fairness" to the student is the range of marks assigned. The student at the bottom is very concerned whether the teacher is one who fails students often. And the student at the top feels that it is "unfair" if there is not a reasonable probability of his getting an A. If an "accurate" or "fair" grade is regarded as that which would be assigned by a group of experts, acting independently of one another, then these questions of mean grade, and of grade dispersion, are very important to such accuracy.

In this way, of course, the machine can be incomparably superior to the human. Both mean grade and grade deviation may be determined entirely on the basis of the expert group, and if desired remain fixed for any group of students for whom the system is applied, regardless of the size of the group. We take such standards for granted with the national standard scores of such instruments as the Scholastic Aptitude Tests, yet ordinarily we despair of trying to achieve the same fairness with any marks administered on a local level. With the introduction of machine essay grading, it appears likely that the parameters of evaluation may be uniformly adjusted to any standards found appropriate.

In the question of accuracy, then, as this quality is ordinarily thought of, the automatic system has some large advantages over the human system, and these are advantages which cannot be easily demonstrated in statistical comparisons. But they should be kept in mind for any thinking about applications.

Using one essay's proxes for another essay's criterion. One way to find out what proxes might have the greatest stability, in terms of measuring important aspects of a student's characteristic writing behaviors, might be to use the proxes from one essay in a multiple regression for another essay by the same students. The reasoning may be obvious: There are certain aspects of student behavior which might influence the human judgment of one essay, but not be much related to the student's long range performance. If we cross the proxes and criterion, then, in the way described, we may be tapping more enduring aspects of writer behavior.

To investigate this question, the proxes from Essay D were used in a new multiple-regression analysis to predict the pooled human judgments for overall quality for Essay C. The resulting MULTR was .62, which as would be expected was a considerable drop from the .71 obtained with Essay C's own proxes. Table IV-12 shows the summary data of interest from this analysis. Column B represents the correlation of the proxes with the criterion, and Column C represents the beta weighting of each prox in the analysis.

Inspection of these columns, and a comparison of them with their counterparts in Table IV-7 and Table IV-3, do not provide any very transparent explanation for the decrement in prediction. A hint may be gained from the slightly lower correlation of essay length with the criterion; students may have more to say on one subject than on another, and this fluency may affect the rater's judgment. And the beta weight for essay length has also dropped markedly (from .32 for Essay D's own criterion, to .21 for Essay C's criterion), which bolsters this suggestion. A comparison of another contributor, standard deviation of word length, shows a similar decrement in beta weight, but an actual increase in the bivariate correlation with the criterion, compared with the Essay D table (IV-3).

In summary of this trial, then, the data are difficult to interpret verbally, but seem to argue that the decrement in multiple correlation may be a reflection of the true difference in student performance across essay topics.

TABLE IV-12

ESSAY D PROXES USED TO PREDICT
AN ESSAY C CRITERION

	A. Proxes	B. Corr. with Criterion	C. Beta wts.
1.	Title present	.03	.07
2.	Av. sentence length	-.01	-.22
3.	Number of paragraphs	.08	.02
4.	Subject-verb openings	-.14	.02
5.	Length of essay in words	.19	.15
6.	Number of parentheses	.11	.05
7.	Number of apostrophes	-.19	-.05
8.	Number of commas	.37	.18
9.	Number of periods	.00	-.09
10.	Number of underlined words	-.03	-.07
11.	Number of dashes	.22	.06
12.	No. colons	.08	.04
13.	No. semicolons	.04	-.00
14.	No. quotation marks	.17	.07
15.	No. exclamation marks	-.07	-.05
16.	No. question marks	-.08	-.11
17.	No. prepositions	.17	.02
18.	No. connective words	.17	.02
19.	No. spelling errors	-.09	-.02
20.	No. relative pronouns	.04	.06
21.	No. subordinating conjs.	-.13	.04
22.	No. common words on Dale	-.44	-.09
23.	No. sents. end punc. pres.	.01	.04
24.	No. declar. sents. type A	.08	.01
25.	No. declar. sents. type B	.06	.02
26.	No. hyphens	.24	.14
27.	No. slashes	-.01	.01
28.	Aver. word length in ltrs.	.45	.10
29.	Stan. dev. of word length	.48	.21
30.	Stan. dev. of sent. length	-.04	.12

Cross-validation with same essays. As we have already suggested, the question of validity is a complicated one. One acceptable form of validity is surely the prediction of future behavior by the same student. But what is often meant is rather the prediction of what expert humans might say about a student's performance. This would become a kind of "concurrent" validity.

In the context of the present project, such concurrent validity would consist of seeing to what degree the machine scorings of the essays would coincide with the human scorings. In one sense, we already know the result to be .71, since this was the discovered multiple correlation for both C and D essays, and represents just what is described: a measure of correlation between the machine scores and the human scores. As we have seen, however, such a coefficient capitalizes upon chance, and should be shrunken statistically. This has also been done, with a resultant shrunken (Wherry) coefficient of .67.

Even the shrunken coefficient is not completely satisfying, however, because of the fact that empirical data often deviate from the assumptions upon which such statistical manipulation is based. Besides, it is desirable to know how the machine algorithm will correlate with the individual judges.

For these reasons, it is most desirable to select randomly among the essays, and generate the weightings from this sub-sample, after which the weightings may be used to assign scores to those essays not included in the multiple-regression analysis. The correlation of these machine scores, may be correlated with the human ratings of these excluded essays, and this new correlation will represent a very appropriate measure of validity. The result of such a procedure is exhibited in Table IV-13.

TABLE IV-13

CROSS-VALIDATION COMPARISON OF

THE COMPUTER WITH FOUR HUMAN JUDGES

(Essay N = 138)

Judges

	A	B	C	D	E
A		51	51	44	57
B	51		53	56	61
C	51	53		48	49
D	44	56	48		59
E	57	61	49	59	

Note: Judge C is the computer. All cells represent correlation coefficients generated by comparing four human judge columns with machine scores on the same essays. The machine scores were those generated from 138 _other_ essays written by _other_ students, chosen at random from the same larger sample.

For this table, the computer-assigned scores, then, were generated from an analysis on 138 of the D essays, which were chosen by random methods from the 276 total sample. The weightings derived from the analysis were then applied to essays written by 138 <u>other</u> students, and the scores so assigned were correlated with the scores assigned by four human judges (Page, 1967a).

This Table IV-13 has often been presented to audiences as the clearest and simplest evidence of the effectiveness of machine strategies, and it has usually been presented without telling the audience which column in fact represents the computer. It is very difficult to guess which one it would be, yet given sufficient time, the occasional sophisticated psychometrician may be able to reason out that Column C is the most probable, and is indeed the computer column. The reason why this is detectable is again characteristic of the difference between man and machine. Surely the machine is not measuring the essay quality in the same way as the man. The machine is surely failing to attend to many of the important syntactic and semantic properties which influence the human judge. But the machine is in one sense more <u>reliable</u> than the human judge, and it is the reliability which gives it away: The coefficients for the machine (Column C) range from only .48 to .53, whereas the coefficients for the human judges (Columns A, B, D, and E) have ranges which are typically three times as large. The machine agrees with the human judges more consistently, then, than they agree with each other!

<u>Practical implications</u>. Although striking, Table IV-13 does not merely represent a simple trick. Rather, it is the clearest analog so far to what might come from a large-scale, machine-based essay evaluation. For example, in a national essay quiz (such as the writing sample occasionally taken by the College Entrance Examina-

tion Board), the procedures would probably be quite similar to those which led to this table. In prior years, a number of essay topics would each be assigned to a fairly small test sample of students. Their writings would be intensively analyzed by expert humans for whatever research and norming purposes might be desired. Then from this pool of tested essay assignments, one stimulus would be chosen to be used across the country on the day of the major test. When collected, these essays would typically not be examined by human judges at all, but would rather be analyzed by computer programs already developed from the test sample. Only a few would subsequently be analyzed by human beings, to check for possible drift in sample, or for historical developments which might have altered the essay topic. In general, the crash scoring, in this hypothetical testing, would typically be entirely mechanical.

Summary. Earlier chapters introduced the rationale, basic design, and initial proxes used in this study, and have presented the computer program used in their measurement. The present chapter has presented some of the findings from the study bearing on the basic questions of the agreement of the human judges with each other, and with machine scores of the same and of different essays. It has furthermore presented information about the proxes: their intercorrelation, their prediction of human judgments, and their reliability across trials. In most important comparisons, the machine scores were found to be practically interchangeable with the human ratings. This finding was most important when various types of validity were analyzed, one based upon prediction of student performance on another occasion, and the other applying measures generated from one set of students to a wholly different set of students. Some comment was also made about inferences from these findings for practical work in the future.

CHAPTER V

PREDICTING A PROFILE OF RATINGS

The last chapter explained in considerable detail the results from the attempts to predict the overall rating of some sets of essays. It was seen that the simulation was indeed very successful, and that the level of success had a number of implications for the future of such work. This chapter considers the more advanced case of simulating an analytic profile of an essay.

If a single overall rating were the only outcome from analysis, it would be a satisfactory substitute for some major tasks today, such as national essay exams (where a single pooled judgment is the usual product of evaluation), or many classroom situations (in which an overworked teacher marks only a letter grade and some redundant comment on a returned essay). Nevertheless, any substantial essay analysis must seek a level of performance nearer to that of the ideal teacher: with a much richer profile of the traits of the writing, so that students (and their instructors) may concentrate differentially on relatively weak skills in the profile; and with more detailed and direct comment about specific patterns or errors in the student's work. This chapter will concentrate on the trait ratings of the essays, and Chapter VIII will give some attention to the detailed and personal comment to the student.

The sample. For the reasons set forth in an earlier chapter, there was no cause for dissatisfaction with the Wisconsin essays (McColly and Remstad, 1963). They did not represent a typical high school student body, but the range was wide, and, with such early strategies, no important interactions with selections were expected. Furthermore, a number of replications had been successfully performed with other essays, with second essays written by the same students, and with a set of essays written by students in Indiana in an unrelated study conducted by Anthony Tovatt and his colleagues at Ball State University under the U. S. Office of Education (and reported elsewhere by Dr. Tovatt). For the phase of work considered in this chapter, therefore, it was decided to continue the analysis, this time working most intensively with Essay C (those based upon the question of whether "the best things in life" were really "free").

The new data needed were judgmental, then, for no evaluative data existed for the Wisconsin essays beyond the simple rating of overall quality. We therefore wished to establish a reliable set of ratings which would constitute a sensible descriptive profile of the strengths and weaknesses most commonly looked for in stylistic judgment. For such a requirement we would need: (1) a set of established and accepted dimensions; (2) a selection of judges who would be representative of qualified English teachers in general; (3) a sufficient number of judgments to overcome the inevitable halo effects, and to establish in truth a meaningful profile.

Just as with the essays, the investigators could afford to be reasonably relaxed about any randomness of selection, so long as judges met the general, personal and professional criteria, because stratification of region, type of school, and a myriad other possible considerations seemed unimportant so far as these particular generalizations of result are concerned. While there are differences among teachers in such dimensions, interaction of such dimensions with the purposes of the study seemed of negligible importance. And what seemed of much greater importance was the control of the rating situation.

The rating session. On July 16, 1966, then, under the principal supervision of Dr. Arthur Daigon, 32 English teachers met at the School of Education of the University of Connecticut for the purpose of grading student compositions for multiple traits. Because this group's judgments of student writing were to represent the evaluations of highly competent professionals, evaluations which would subsequently be simulated by a computer, selection of participants (setting randomness aside) was done with considerable care.

Ten chairmen of English departments in Connecticut secondary schools were invited to participate with teachers whom they could recommend as having special competence in the grading of student compositions, and who had at least three years of teaching experience. The department chairmen were also requested to give first preference to those skilled teachers who possessed master's degrees.

Of the 32 teachers who participated, 10 (31 %) were department chairman and 28 (87 %) possessed M.A. degrees. The mean number of years of English teaching experience was 12.9 years, the median, 10 years.

Before the grading session began, the teachers were welcomed and acquainted with their task. Each would grade 64 compositions, assigning separate grades on a 5 point scale for each of 5 traits designated as "ideas or content", "organization", "style", "mechanics", and "creativity". Each English teacher-judge received both written and oral instructions relating to identification and scoring of the traits. Samples of a "good" composition and of a "poor" composition were distributed and considered in order to demonstrate how the traits could be scored and to suggest a range of possible response.

Eight 30-minute time periods were established. During each period each judge evaluated 8 compositions, which allowed about 3-1/2 minutes for the multiple trait judging of each composition. These arrangements permitted 8 judgments for each of 5 traits in each of the total of 256 compositions.

The assignment of essay to teacher and period was a formidable task, which required the computer. The problem was manifold: each essay could be assigned only once in each of the eight periods, so that there would be no important period effect hurting the essay evaluations, and so that we would not need multiple copies of the essays. Yet no essay could be given to any judge more than once. And it was also desirable to randomize the order of presentation within a period. These problems are rather easily solved if groups of eight essays are kept together, but such a procedure would obviously distort the evaluation of an essay in unknown ways. There is another major problem, in that it is easy to continue random assignment up to the last period, and find unresolvable conflicts of assignment, requiring branching back to some earlier point. But eventually, with intensive work, the problem was solved and made completely automatic.

The mechanics of the rating day were not a trivial problem, however, since we needed six graduate assistants performing the reassignments. Luckily, though, they could use punched and interpreted assignment cards, a by-product of the computerized assignment program, which also served as mark-sense rating records for later analysis.

The rating criteria. In choosing the traits for rating, we desired well-established dimensions of writing quality. One of the most helpful documents was an eight-scale evaluation designed by Paul Diederich, and used at the Educational Testing Service ("Definitions of Ratings on the ETS Composition Scale" -- no date). Figure V-1 shows our adaptation of such suggestions.

Naturally, there are large differences among raters in their evaluations of the same essays. Since each rater read 64 of the essays, and since this number represents one fourth of the 256 used in the design, the probability of any one essay being read by any judge is clearly 1/4, and we might expect that, of the 64 essays read by judge A, about 1/4, or about 16, would be read by judge B. Clearly, with such small N's, we would not expect very secure estimates of the population agreement between two judges, but would rather expect a large random error.

Table V-1 shows the intercorrelations among all 32 judges. The correlations are based upon the "total" scores, which were the average of the five trait ratings given by any judge to an essay. As can be seen, the median judge intercorrelation hovers around .5 for these total scores.

The judges were instructed, as is clear in Figure V-1, to balance their ratings into a certain distribution, approximately normal, and it would be expected that their means and standard deviations would therefore be approximately equal. Table V-2 shows that this is indeed the case. Since 5 represents the best rating, and 1 represents the worst, the means are all seen to deviate from the expected 2.5 in a slightly generous direction. The nebulous "ideas" or "content" is the most tolerantly graded, with "organization" a second place. "Mechanics" has a middle position of severity of marking, and has decidedly the largest standard deviation of any trait. Teachers were thus more decisive about mechanics and, as we shall see later, they agreed more with each other about mechanics than they did about the other dimensions of essay quality.

FIGURE V-1

CRITERIA FOR RATING THE ESSAYS

I. Definitions of the basic traits to be rated.

A. Ideas or Content. The quantity and quality of the materials used to cover the subject.

B. Organization. The relationship between the parts of the paper and the whole.

C. Style. The use of language above and beyond the problems of mechanics.

D. Mechanics. Spelling, grammar, usage, punctuation, capitalization, numbers.

E. Creativity. The degree to which the paper finds a new, unexpected, yet fruitful way to approach the subject, to combine ideas, and to utilize language. An over-all trait.

II. Guides for rating the basic traits.

A. IDEAS OR CONTENT

High. The student covers the materials that the topic and plan of attack clearly call for. His understanding of the subject is good and he uses clear definitions. He has the ability to see the topic in a broader perspective than do the other students in his group, that is, he brings a broader experience to the topic.

Middle. The ideas are appropriate, but conventional and few in number. Some aspects of the topic are left out. The writer does not seem to have a well-stocked mind.

Low. The student omits many important aspects of the topic. He seems to have no store of knowledge to bring to bear on the topic and consequently repeats a few simple ideas over and over again.

B. ORGANIZATION

High. The student has a definite plan for discussing the assigned topic. If he is arguing for or against an idea, he presents relevant reasons in an effective order. If he is describing something, he does so according to some scheme (top to bottom, order of importance, order of complexity, etc.) If the student is explaining a concept or process he uses a coherent plan of analysis, or definition, or illustration. The student has a good sense of what is relevant to his plan and avoids repetition. He shows a sense of proportion in treating the various parts of his essay.

FIGURE V-1 (cont.)

Middle. The student shifts his plan of discussion, or introduces irrelevant material, or spends too much time on unimportant things, or repeats himself. He develops the assigned topic by free association (what comes to my mind when I think of Hawaii?) rather than by working toward a definite purpose.

Low. The student does not seem to have given any thought to what he intended to say before he started to write. He offers no plan of discussion. The paper seems to start in one direction, then another, then another, until the reader is lost. The main points are not clearly separated from one another, and they come in a random order.

C. STYLE

(There are many aspects of style that may enter into a rating--individuality, vividness, elegance, etc. However, for the purposes of this experiment we are interested in three stylistic traits only--clarity, variation, and range of linguistic resources.)

High. The student uses language in a way that makes comprehension of the paper easy. He uses appropriate words in their normal sense. He puts the words in their normal order. He is careful to signal his transitions. He avoids ambiguity and he does not frustrate the reader's expectations. At the same time the student avoids monotonous repetitions of words, phrases, and sentence structures. Finally, he reveals a command of a good range of linguistic resources. His vocabulary is good, he uses parallel structures, he makes subtle use of subordination, and so on.

Middle. The student occasionally brings the reader up short by choosing a bizarre, inappropriate word or phrase, or by introducing a distracting metaphor, or by misplacing a modifying phrase or clause, or by making unexpected transitions. The repetitions of words, phrases, and sentence structures become monotonous. The resources of language are limited. The writer is addicted to tired old phrases and hackneyed expressions.

Low. Vague use of words. Ambiguous references. Awkward constructions. Childish vocabulary and sentence structure.

FIGURE V-1 (cont.)

D. MECHANICS

<u>High</u>. The sentence structure is usually correct, even in varied and complicated sentence patterns. No violation of established spelling rules. Even the hard words are usually spelled correctly. No serious violations of the rules of punctuation, capitals, abbreviations, and numbers.

<u>Middle</u>. An occasional syntax problem. Hard words are occasionally mispelled. Some violations of the rules concerning punctuation, etc.

<u>Low</u>. The student borders on the illiterate.

E. CREATIVITY

<u>High</u>. The student surprises us with a new and fruitful way of looking at the problem. He brings to bear new data in treating the topic. He finds a fresh and interesting way of using language that illuminates his ideas.

<u>Middle</u>. The student thinks of the expected things. He treats them in a way that most people would treat them. He makes use of ordinary expressions and sentence structures.

<u>Low</u>. The student works with cliches of thought and expression. Does not go beyond the most superficial treatment of the subject. Repeats formulas without really grasping their meaning.

Try for the following overall balance

RATING	
5	<u>TOP</u> 7% or so. About 2 of each 16 essays.
4	<u>NEXT</u> TOP 25%. About 4 of each 16 essays.
3	<u>MIDDLE</u> 35%. About 6 of each 16 essays.
2	Next <u>BOTTOM</u> 25%. About 4 of each 16.
1	<u>BOTTOM</u> 7% or so. About 2 of each 16.

TABLE V – 1

INTERCORRELATION MATRIX FOR 32 RATERS FOR TOTAL SCORES

Correlation	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
1	*	.33	.54	.33	.65	.75	.43	.56	.05	.68	.32	.20	.88	.78	.60	.49	.55	.71	.36	.75	.26	.49	.81	.75	.84	.42	-.02	.26	.51	.66	.29	.45
2		*	.50	.36	.15	.24	.74	-.16	.24	.39	.28	.59	.26	.20	.40	-.13	.34	.70	.48	.41	.35	.64	-.00	.61	.37	.65	.51	.47	.53	.00	.43	.42
3			*	.62	.79	.64	.59	.48	.39	.70	.38	.56	.53	.66	.49	.50	.55	.39	.36	.40	.60	.57	-.06	.68	.67	.40	.39	.74	.82	.55	.27	.38
4				*	.76	.43	.42	.46	.65	.88	.43	.64	.58	.52	.19	.27	.81	-.08	.57	.48	.36	.31	.65	.35	.06	.38	.48	-.14	.49	.13	.80	.57
5					*	.59	.20	.23	.75	.70	.81	.26	.74	.83	.76	-.01	.43	.04	.89	.33	.56	.41	.69	.72	.44	-.09	.26	.35	.68	.31	.49	.71
6						*	.45	.20	.45	.49	.58	.73	.66	.32	.25	.22	.41	.46	.32	.53	.16	.50	.61	.62	.66	.64	.55	.85	.57	.79	.67	.51
7							*	.52	.51	.22	.52	.26	.62	.23	.66	.08	.30	.62	.85	.05	.56	.40	.51	.87	.47	.49	.46	.53	.84	.53	.47	.49
8								*	.36	.70	.24	.03	.81	.43	-.02	.42	.71	.21	.66	.20	.36	.39	.22	.25	.11	.18	.18	.28	.64	.50	.38	.33
9									*	.47	.22	.78	.52	.44	.51	.36	.23	.48	.26	.14	.15	.55	.59	.42	.48	.14	.67	-.04	.76	.06	.21	.58
10										*	.56	.21	.73	.10	.25	.23	.65	.51	.79	.79	.50	.67	.49	.49	.65	.51	.09	.56	.78	.07	.54	.16
11											*	.01	.32	.27	.70	.34	.54	.15	.60	.42	.95	.45	-.01	.55	.77	.28	.32	.77	.52	.28	.31	.26
12												*	.63	.42	.60	.62	.52	.67	.71	.17	-.01	.60	.27	.72	.34	.45	.45	.39	.25	.08	.44	.52
13													*	.62	.51	.37	.66	.58	.89	.35	.35	.38	.86	.82	.78	.59	.06	.39	.72	.43	.51	.72
14														*	.77	.87	.46	.12	.80	-.07	.22	.40	.71	.21	.75	.28	.55	.11	.77	.64	.42	.49
15															*	.16	.43	.07	.89	.74	.32	.72	.06	.67	.64	.63	.18	.10	.22	.05	.09	.04
16																*	.64	.56	-.03	.07	.15	.03	.71	-.09	.49	.58	.34	.44	.36	.19	.27	.08
17																	*	.30	.52	.44	.36	.69	.65	.68	.50	.59	.52	.64	.49	.15	.72	.82
18																		*	.44	.61	.33	.27	.46	.28	.66	.19	-.13	.29	.38	.46	.73	.18
19																			*	.64	.57	.62	.52	.54	.72	.40	.15	.57	.53	.42	.80	.75
20																				*	.26	.42	.44	.71	.38	.25	.27	.51	.62	.32	.06	.54
21																					*	.26	.08	.25	.38	.06	.56	.22	.33	.26	.59	.09
22																						*	.69	.74	.30	.79	-.43	.63	.66	.45	.38	.88
23																							*	.40	.76	.73	.17	.30	.47	.14	.08	.21
24																								*	.64	.52	.56	.69	.64	.01	.03	.65
25																									*	.67	.23	.09	.69	.37	.69	.79
26																										*	-.05	.56	.55	.20	.61	.05
27																											*	-.05	.51	.72	.44	.56
28																												*	.55	.41	.35	.34
29																													*	.70	.33	.34
30																														*	.33	.08
31																															*	.28
32																																*

TABLE V-2

MEANS AND STANDARD DEVIATIONS OF THE

FIVE TRAITS AND THEIR AVERAGE

Trait	Mean	Standard Deviation
1. Ideas	3.068	.640
2. Organization	2.950	.675
3. Style	2.827	.619
4. Mechanics	2.869	.771
5. Creativity	2.833	.641
6. Trait average	2.909	.610

Contribution of the proxes. It is of interest to see how each prox contributed to the prediction of the five various traits, and this information is contained in the next five tables (V-3 to V-7). The information for the average of the five traits is contained in Table V-8. The first column has the number of the proxes and, in the first of these tables (V-3) the name of the prox as well. Column B has the correlation with the criterion for each prox. Column C has the B-weights for each prox. And Column D has the computed t-values for each prox.

Column C, which has the B-weights for each prox, should not be confused with the Beta weights given, for example, in Table IV-3 of the last chapter. The B weight is the coefficient which is actually used, together with the raw prox score, to optimize the predictive value of any prox in an applied situation. In other words, given two proxes of the same Beta coefficients, the one having a larger standard deviation will have a smaller B-weight. While these B-weights may not be compared directly with those Beta coefficients given in the last chapter, they may be compared with the corresponding B-weights of the other traits given in this chapter, though any such comparison would be simply monotonic, and differences could not be easily compared between proxes.

The relative contribution of each prox may be inferred from the t-values in Column D. The absolute values of these t's are monotonically related to the rank order of contribution of the proxes to the prediction. For example, Table V-3 shows that the highest contribution was made by fifth prox, showing a t-value of 6.38, far ahead of any other. When it is considered that Prox #5 is "length of essay," and that the trin is "ideas or content," we are struck by the obviousness of the relation. The more words used, the more content the essay is believed to have. For

TABLE V-3

PROX CONTRIBUTION TO THE PREDICTION OF
IDEAS OR CONTENT
(N = 256)

	A. Proxes	B. Corr. with Criterion	C. B wts.	D. F-value
1.	Title present	.01	.14973	1.37
2.	Av. sentence length	-.07	-.00341	-1.73
3.	Number of paragraphs	.14	-.03357	-1.79
4.	Subject-verb openings	-.19	-.00326	-1.38
5.	Length of essay in words	.37	.00205	6.38
6.	Number of parentheses	-.04	-.00875	-1.23
7.	Number of apostrophes	-.11	-.00640	-1.73
8.	Number of commas	.37	.00601	3.73
9.	Number of periods	.03	-.00007	-0.01
10.	Number of underlined words	-.04	-.01053	-1.14
11.	Number of dashes	.36	.03283	4.63
12.	No. colons	.07	.02032	1.03
13.	No. semicolons	.13	.00964	0.90
14.	No. quotation marks	.15	.00215	1.25
15.	No. exclamation marks	-.05	-.00272	-1.26
16.	No. question marks	.08	-.00121	-0.56
17.	No. prepositions	.17	.02926	1.54
18.	No. connective words	.09	.00797	0.12
19.	No. spelling errors	-.10	-.05672	-0.55
20.	No. relative pronouns	-.07	.03506	1.09
21.	No. subordinating conjs.	-.09	.02955	0.88
22.	No. common words on Dale	-.34	-.00270	-0.22
23.	No. sents. end punc. pres.	.17	.03073	1.33
24.	No. declar. sents. type A	-.00	-.01298	-0.60
25.	No. declar. sents. type B	.04	-.01483	-0.52
26.	No. hyphens	.23	.00717	0.99
27.	No. slashes	.05	.05393	0.66
28.	Aver. word length in ltrs.	.34	-.00094	-0.34
29.	Stan. dev. of word length	.43	.00921	3.17
30.	Stan. dev. of sent. length	.11	.00200	1.38
	Intercept constant		-1.01123	
	Multiple correlation		0.72301	
	Std. error of estimate		0.47093	

F multr = 8.21

TABLE V-4

PROX CONTRIBUTION TO THE PREDICTION OF ORGANIZATION
(N = 256)

A. Proxes	B. Corr. with Criterion	C. B wts.	D. F-value
1.	.01	.10786	0.82
2.	-.11	-.00229	-0.96
3.	.10	-.01289	-0.57
4.	-.20	-.00343	-1.21
5.	.23	.00124	3.20
6.	-.08	-.01417	-1.65
7.	-.08	-.00253	-0.57
8.	.31	.00515	2.65
9.	.08	.00399	0.53
10.	-.06	-.01267	-1.14
11.	.26	.02136	2.50
12.	.07	.02091	0.88
13.	.10	.00449	0.35
14.	.18	.00035	2.44
15.	-.03	-.00287	-1.11
16.	.08	-.00199	-0.78
17.	.12	.01855	0.81
18.	.13	.05610	0.72
19.	-.16	-.24534	-1.98
20.	-.07	.03517	0.91
21.	-.10	.01539	0.38
22.	-.33	-.00580	-0.39
23.	.17	.04423	1.59
24.	-.01	.02481	-0.95
25.	.06	-.02201	-0.64
26.	.20	.00527	0.60
27.	.02	.01598	0.16
28.	.33	.00082	0.25
29.	.38	.00695	1.99
30.	.05	.00118	0.68

Intercept Constant	-1.29180
Multiple correlation	0.61455
Std. error of estimate	0.56709

F multr. = 4.55

TABLE V-5

PROX CONTRIBUTION TO THE PREDICTION OF
STYLE
(N = 256)

A. Proxes	B. Corr. with Criterion	C. B wts.	D. F-value
1.	-.02	.10535	1.01
2.	-.12	-.00385	-2.04
3.	.07	-.03715	-2.08
4.	-.17	.00067	0.30
5.	.23	.00123	3.99
6.	-.02	-.00505	-0.74
7.	-.07	-.00254	-0.72
8.	.41	.00624	4.05
9.	.08	.00316	0.53
10.	-.05	-.01147	-1.30
11.	.32	.02689	3.97
12.	.06	-.00236	-0.13
13.	.13	.00880	0.86
14.	.16	.00019	1.64
15.	-.01	-.00329	-1.60
16.	.11	-.00276	-1.35
17.	.19	.00127	2.28
18.	.15	.11271	1.82
19.	-.15	-.20081	-2.04
20.	-.06	.05119	1.66
21.	-.09	.04533	1.41
22.	-.39	-.00476	-0.40
23.	.18	.04894	2.22
24.	-.04	-.03366	-1.62
25.	.05	-.02750	-1.00
26.	.32	.01625	2.34
27.	.05	.05503	0.70
28.	.40	-.00044	-0.17
29.	.47	.00906	3.26
30.	.11	.00315	2.28

Intercept constant -1.33964
Multiple correlation 0.72951
Std. error of estimate 0.45042
F multr. = 8.53

TABLE V-6

PROX CONTRIBUTION TO THE PREDICTION OF MECHANICS
(N = 256)

A. Proxes	B. Corr. with Criterion	C. B wts.	D. F-value
1.	-.00	.11647	0.83
2.	-.17	-.00258	-1.02
3.	-.01	-.02655	-1.11
4.	-.08	.00380	1.26
5.	.06	.00053	1.29
6.	-.02	-.00483	-0.53
7.	-.07	-.00234	-0.49
8.	.29	.00579	2.80
9.	.16	.00684	0.85
10.	-.06	-.00745	-0.63
11.	.23	.02192	2.42
12.	.12	.03514	1.39
13.	.10	.01353	0.99
14.	.09	.00017	1.07
15.	-.05	.00106	0.39
16.	.04	.00288	1.05
17.	.17	.06407	2.64
18.	.14	.17301	2.09
19.	-.30	-.68565	-5.21
20.	-.08	.01675	0.41
21.	-.06	.04182	0.97
22.	-.28	.03763	2.36
23.	.21	.00580	0.20
24.	.07	.02471	0.89
25.	-.01	.00433	0.12
26.	.25	.02133	2.30
27.	.08	.10975	1.05
28.	.39	.00535	1.54
29.	.42	.01128	3.04
30.	.00	.00136	0.74

Intercept constant	-9.53911
Multiple correlation	0.67796
Std. error of estimate	0.60314

F multr. = 6.38

TABLE V-7

PROX CONTRIBUTION TO THE PREDICTION OF
CREATIVITY
(N = 256)

A. Proxes	B. Corr. with Criterion	C. B wts.	D. F-value
1.	-.02	.13550	1.21
2.	-.12	-.00037	-0.19
3.	.09	-.05943	-3.11
4.	-.17	-.00130	-0.54
5.	.34	.00208	6.36
6.	-.08	-.01716	-2.36
7.	.00	-.00026	-0.07
8.	.39	.00648	3.94
9.	.11	.01684	2.64
10.	-.06	-.01797	-1.91
11.	.29	.02680	3.70
12.	.09	.01817	0.90
13.	.16	.02121	1.94
14.	.12	.00015	1.18
15.	.00	-.00363	-1.65
16.	.10	-.00295	-1.35
17.	.13	.02480	1.28
18.	.07	.02156	0.33
19.	-.11	-.09013	-0.86
20.	-.07	.04230	1.29
21.	-.08	.06215	1.81
22.	-.32	-.01342	-1.06
23.	.15	.03678	1.56
24.	-.05	-.03993	-1.80
25.	.04	-.04165	-1.42
26.	.30	.01846	2.49
27.	.05	.06153	0.74
28.	.28	-.00322	-1.16
29.	.36	.00820	2.77
30.	.10	.00232	1.57

Intercept constant	1.21571
Multiple correlation	0.70938
Std. error of estimate	0.48074

F multr = 7.60

TABLE V-8

PROX CONTRIBUTION TO THE PREDICTION OF
AVERAGED RATING ACROSS 5 TRAITS
(N = 256)

A. Proxes	B. Corr. with Criterion	C. B wts.	D. F-value
1.	-.00	.12299	1.18
2.	-.13	-.00250	-1.33
3.	.08	-.03392	-1.90
4.	-.18	-.00071	-0.31
5.	.26	.00143	4.65
6.	-.05	-.00999	-1.47
7.	-.07	-.00281	-0.80
8.	.38	.00593	3.86
9.	.10	.00615	1.03
10.	-.06	-.01202	-1.37
11.	.32	.02596	3.84
12.	.09	.01844	0.98
13.	.14	.01154	1.13
14.	.15	.00020	1.75
15.	-.03	-.00229	-1.11
16.	.09	-.00121	-0.59
17.	.17	.03559	1.97
18.	.13	.07427	1.20
19.	-.19	-.25574	-2.61
20.	-.08	.03609	1.17
21.	-.09	.03885	1.21
22.	-.36	.00219	0.18
23.	.20	.03330	1.51
24.	-.01	-.01733	-0.83
25.	.04	-.02033	-0.74
26.	.28	.01370	1.98
27.	.06	.05924	0.76
28.	.38	.00031	0.12
29.	.45	.00894	3.23
30.	.08	.00200	1.45

Intercept constant -2.39336

Multiple correlation 0.72145

Std. error of estimate 0.44944

F multr = 8.14

no other trait is essay length quite so dominant, though it plays an almost equal role for Table V-7, where "creativity" is the trin. When we consider how often creativity is measured by tests of fluency and fecundity, we are again struck by the obviousness of the relation. Similar comparisons can be made for other traits and other proxes.

If we rank order the top five contributors for each trait separately, the results are interesting: For the trait of <u>ideas</u>, we find the proxes, according to the absolute value of Column D, to contribute in the following order:

1st) length of essay
2nd) frequency of dashes
3rd) frequency of commas
4th) standard deviation of word length
5th) number of paragraphs

For the trait of <u>organization</u>, the order of contributory proxes is:

1st) length of essay
2nd) frequency of commas
3rd) frequency of dashes
4th) frequency of quotation marks
5th) standard deviation of word length

For <u>style</u>

1st) frequency of commas
2nd) length of essay
3rd) frequency of dashes
4th) standard deviation of word length
5th) frequency of hyphen

For mechanics:

1st) spelling errors

2nd) standard deviation of word length

3rd) proportion of prepositions

4th) frequency of commas

5th) frequency of dashes

For creativity:

1st) length of essay

2nd) frequency of commas

3rd) frequency of dashes

4th) number of paragraphs

5th) standard deviation of word length

And for the average of all five traits, calculated for each essay, the order is as follows:

1st) length of essay

2nd) frequency of commas

3rd) frequency of dashes

4th) standard deviation of word length

5th) spelling errors

There is obvious noise in any comparisons of such listings. In the first place, there is random error, which is considerably higher in calculating Beta weights, or these similar multivariate t-values, than in calculating bivariate relationships. In the second place, in the trins themselves there is a high degree of halo effect, as will be shown soon. We would expect the first of these problems to be exhibited in rather wild and unexplained loadings, that would not necessarily be replicated in cross-validations. We would expect the second problem to be evident in the occurrence of some common proxes in all lists, as here we see word length to be an important correlate of all traits, and commas to be another.

Nevertheless, there are ways in which the differences among these rankings are intuitively pleasing. Length of essay is of first importance for three traits, and second for a fourth. But on one list, essay length does not occur at all, and this one is the list for mechanics. On the other hand, for mechanics we find the only inclusion of spelling errors. The evaluation of mechanics is clearly a rather negative thing, in which mistakes count against the student, and it seems better for a student to be short and safe, than to be fluent.

In summary, these tables furnish many interesting comparisons for differential study of the contributions of the proxes to the central dimensions of ratings which were studied here. While there is considerable overlap of the important proxes, there is also some difference in weighting which increases the accuracy of prediction.

The uniqueness of the traits. A constant danger in multi-trait ratings is that they may reflect little more than some general halo effect, and that the presumed differential traits will really not be meaningful. This danger is one reason for having eight judgments for each essay, since it was predicted that the halo would be extremely large. And the evidence we have already seen, showing the relative contribution of the proxes to the prediction of the traits, supports this suspicion of a large halo effect.

This halo is demonstrated in Table V-9, which shows the intercorrelations among the judged traits of the essays, as rated by eight teachers for each essay. From this table it is clear that mechanics is the most maverick trait, having little to do with ideas, organization, or creativity, but considerably more to do with style. We find a very large halo, or tendency for ratings to agree with each

TABLE V-9

INTERCORRELATIONS OF THE TRAITS
OF JUDGED ESSAYS
(N = 256)

TRAIT	1	2	3	4	5	6
1. Ideas		.86	.86	.68	.89	.93
2. Organization	.86		.82	.69	.82	.91
3. Style	.86	.82		.83	.86	.95
4. Mechanics	.68	.69	.83		.65	.85
5. Creativity	.89	.82	.86	.65		.92
6. All traits	.93	.91	.95	.85	.92	

other. It may be noticed that, because of the interdependence of these ratings, and their mode of assignment by the judges, the intercorrelations here are in some cases actually larger than the true reliability of the group ratings. Some reflection will show how this could be: once a general level of rating is assigned, reliable or not, it will carry all the traits along together with it. The variable "all traits" on Table V-9 refers to the average of the five traits for an essay, counted as equally important. Naturally, the large correlations shown between "all traits" and the individual traits are inflated by a part-whole relationship.

To test the uniqueness of the traits, James J. Roberge and others performed the analysis of variance shown in Table V-10. There is of course a huge variance between essays, and we also find a large variance between traits (explained by the mean differences we saw in Table V-2). What is important in this Table V-10 is the significant trait-by-essay interaction, which demonstrates that there is a reasonably reliable profile displayed, and that indeed there is some "validity" in the different ratings.

It would be possible, of course, to extract the halo, and to work with the residual, and unique, trait variance for various prediction purposes. We chose not to do this for two reasons: In the first place, we would need considerably more raters for each essay, since the residual trait variance, after the halo was subtracted, would be far less reliable than the original rating, and would make a much less secure goal to simulate. In the second place, and more importantly, we were interested in simulating the real ratings actually given for a certain trait by real human judges with appropriate expertise. And when this is a primary goal, then the halo behavior is an appropriate part of the simulation target, whether "pure" or not.

TABLE V - 10

TRAIT BY ESSAY INTERACTION

Source	SS	df	MS	F
Between judgments	8,230.305	2,047		
Between essays	3,791.293	255	14.868	6.002
Error between	4,439.012	1,792	2.477	
Within judgments	3,564.414	8,192		
Between traits	84.212	4	21.053	56.412
Trait x essay	805.089	1,020	.789	2.115
Error within	2,675.113	7,168	.373	
Total	11,794.719	10,239		

*This table is based upon essay evaluation of July 1966, during which each of 256 essays was judged by eight different judges during eight different periods.

Judge viewpoints. One effort within this project to improve the predictability of judgments was undertaken by Herbert Garber and Robert Shostak (1967), and reported at the Annual Meeting of the American Psychological Association. Their work aimed at raising the multiple R, but by purifying the trin, rather than by optimizing the proxes. Much of this section is quoted or paraphrased from their presentation.

It is well known that one source of error variance in a correlation study is the unreliability of the criterion. Indeed, working under the assumption of a multivariate normal population, it seems reasonable that in any random sample, errors ought to be distributed with equal frequency in all variables, including the criterion (Hays, 1963, p. 573).

The inspiration for this particular approach came, in part, from work reported in _Educational and Psychological Measurement_, volumes 26 and 27, by Naylor and Wherry, and also from an article by Jackson and Messick (1961). The first two investigators used a factor-analytic approach to do what they call "capture rater policy". They used as subjects Air Force supervisory personnel. The latter pair described a similar technique for use in studying social perception of personal status. One procedure which is related to the one of Garber and Shostak is also reported by Christal (1963).

The departure in the present section was first simply to find, from among the 32 reader-graders in Project Essay Grade, those clusters of readers who tended to agree with one another no matter what their policy. In this case, their revealed agreement would emerge from factor-analyzing judges, not essay grades. The next step after identifying a "clear" cluster of judges was to use their individual unpooled grades as the criterion in a new multiple correlation compution to see if the multiple coefficient would rise. If it would, then the demonstration had worked as hoped. A set of judges had been revealed by the analysis who were to a <u>larger</u> extent "predictable" from a fairly mechanical, computer-executed count of what we have labelled "proxes".

What Garber and Shostak wanted was to have a factor analysis done on raters, and since as few as five essays, in one case, had been read in common by a pair of raters, it seemed impractical to proceed to the computation of an intercorrelation matrix from a data matrix 3/4's empty. However, Dieter Paulus arranged to have the incomplete 256 by 32 score matrix processed at Cornell by Larry Wightman. The Cornell program inserted appropriate correlation values based on the varying numbers of observations available for each essay. On the average, 16 were present. Thus prepared, the Cornell computer next processed the judge score <u>intercorrelations</u> and computed factor matrices both by the components and factor analysis procedures. The eleven-factor matrix from the latter computation produced about three or four fairly clear clusters comprised of about as many individual judges. By "clear" is meant a positive loading of over .65 on a single factor and no other positive loadings greater than .43. Negative loadings were ignored.

One cluster comprised of four judges which met the above criteria was selected by inspection. For each of these judges, 64 essays and the rating he gave to each were gathered and then served as input to the next step in the project. A modification of the IBM Scientific Subroutine Program (SSP) for System 360 on multiple regression was used at the University of Connecticut to compute an overall multiple R from the four readers and their essays' prox scores. The result was a coefficient of .65. At first blush this looks as if it were a disappointing outcome until one is reminded that the multiple prediction is based in this instance on an unpooled, unweighted criterion so that any disagreement among the four readers and any dissimilarity in the proxes of the four more or less unique essay sets they read both contribute to a lowered overall reliability. For one must remember that among these four highly correlated judges who were finally picked to be a test case of a refined criterion, most pairs had read no more than 1/4 of their essays in common. Therefore, to get some reasonable basis of comparison, four other judges were selected by a random procedure and their data were treated in exactly the same way. This time, a multiple coefficient of only .545 resulted. Comparing the respective coefficients of multiple determination, .420 and .297, we see that a bit over 12% more of the total variance in the criterion scores has been accounted for when using the selected judges. Putting it in terms of forecasting improvement over the "random four" judges we realize that using the technique here described we have a 40% improvement over "chance".

Let us restate what was done by Garber and Shostak. There was a fixed sample of high school essays. A multiple correlation coefficient of over .71 was computed when an averaged rating obtained from eight randomly-drawn readers from a 32-reader pool was utilized as the criterion; and 30 approximations to writing ability such as length of essay, number of commas, use of uncommon words, and standard deviation of word length served as the predictors. Next, a factor analysis based on the unequal numbers of observations for the 496 pairs of judges in the 32-judge pool was calculated. From this analysis one of several sets of variables (really judges), which were more or less clearly identifiable by the simple structure criterion, were selected as new criteria for a multiple correlation computation. However, this time the prediction would be much more stringent. There could be no approximation to a "true" grade for each essay in the usual test-theory sense of a mean score from repeated sampling. Instead, the error term would contain increased variance from two sources and to an unknown degree. These error components were, on one hand, from the relative lack of overlap in the actual essays read mutually by four judges, as contrasted with the higher number of "same" essays that eight judges from a 256-essay "population" had read. The other error source was the lack of the beneficial effects due to cancelling out random errors which occurred when eight ratings were averaged for each essay.

To get an estimation as to what had been gained from this method of judge selection, a comparison was made with a random selection of four other judges from the remaining 28. Nearly 40% more predictability was found to be the estimated gain. A sampling distribution of multiple R's could have been gathered, and thus a crude sort of significance approximation calculated, for the R obtained on the selected cluster and on the other untried clusters. However, one would have serious misgivings about any generalizability of such findings to other samples from the same population of essay writers and graders.

What are some implications from this study? In the words of Garber and Shostak:

> First, it has been shown empirically that, by this technique, one specific small gain may be made toward the goal of increasing the multiple R in essay grading by computer through selecting of criteria on the basis of clusters of consistent viewpoint among a random sample of readers.
>
> Second, by some technique like stepwise regression using such identified clusters for criteria, knowledge may emerge about the essay evaluation process itself.

And, third, as Davis (1965) suggested in his critique of Project Essay Grade, we may by the route marked out here avoid wiring students to an "easy mass standard" writing style since, instead of exposing merely a simplicity in human writing behavior, we can begin to uncover some sources of dissonance and raucous rumblings among the lions who rule the teaching of effective composition.

The possibility outlined here has not yet been capitalized upon for the production of higher multiple-R coefficients, but it may be regarded as a tool for future investigation. It may in the future be extended to trait analysis as well as an overall dimension, if it proves feasible in later study.

Trait prediction by machines. For our present stage of development of a new discipline, perhaps the best comparisons may still be those of the human expert compared with the machine. As we have seen, eight expert judges, randomly selected from a qualified panel of 32 such judges, read every essay and evaluated it for ideas, organization, style, mechanics, and creativity. We were particularly interested in including the last named, because of the common objections encountered to this sort of work by some teachers in the humanities.

From the beginning, humanists have often miscalculated the difficulties in essay analysis, and imagined that specific criticisms of punctuation and usage might be easy to program for the computer, but that global measures such as overall quality, style, or creativity would be virtually impossible. In one sense, quite the reverse is true: We have had prompt success in actuarially simulating the ratings of these subjective traits, as all of our data reported so far would suggest. Yet a really sound decision about the correctness of usage of a comma, or the agreement of subject and verb, is a prolem which presumes a great amount of analysis, and some of the necessary background routines have not yet been programmed for any project anywhere. We shall discuss these problems later.

Surely, from this humanistic point of view, the most challenging problem of all would be to measure creativity, since by such reasoning a creative work, or original work, is by definition unlike the others, and is unique; and therefore it requires a recognition procedure which could not be programmed in advance.

Obviously, the first step is to remove the problem from the humanistic viewpoint and put it within a viewpoint susceptible to behavioral analysis. In order to do this, we must ask: How is creativity recognized? How do we know when we have achieved it? The only possible answer seems to be that a work is creative when people <u>say</u> it is creative; there is no evaluative procedure above human judgment for deciding whether something is imaginative, or original.

But once again, we must appeal to behavioral science. If we use <u>one</u> judge (and the humanist, when pressed, will often designate himself as sole arbiter), then we have a very uncertain criterion. We do not know to what extent the evaluations made by this judge will correspond with the "true" creativity in the work. Therefore, we must ask other judges to assess the work independently of each other and of the first one, and we must regard their judgments, in absence of evidence to the contrary, as equally valid, if they are equally "authorities" in such matters (however qualifications might be established). We still do not know how well these judgments correspond with the "true" creativity in the work, but at least we can ascertain how well they correspond with each other.

We must, in the end, assume that the population of all such expert ratings would indeed represent our best estimate of any such "true" creativity. Not to admit this would leave us in hopeless solipsism. And when we do admit such a criterion, and when such ratings are made of large numbersof essays, each of which may or may not possess "creativity," we are led to additional discoveries about the trait. And these discoveries are quite contrary to the usual humanist way of thinking.

For the distribution of creativity turns out to be approximately normal, and approximately continuous. And it has (as we note in Table V-2) a standard deviation of .641 rating points, which is just in the middle of the five traits. That is, it does not appear to be a purely "qualitative" trait, which may at once be recognized for its prescenceor absence. Furthermore, to emphasize the apparent continuous quality of the trait, the reliability of human judgment of creativity was the lowest for any of the five traits, as will be seen presently.

In short, then, there is every reason to regard "creativity" as a criterion rating like any other. And to regard in the same way originality, imagination, and other near synonyms used or implied by the instructions to the raters, shown in Figure V-1. Of course, the distribution is in part a result of the instructions regarding such distributions, but there was no apparent tendency by the teachers to force it into a yes-no pattern.

The data for all five traits, then, are shown in Table V-11, which may represent the most complete statement yet about the comparative success of the basic proxes so far presented. Column A of course lists the traits by title. Column B shows us the reliability of the pooled sum of eight independent judges, calculated for each trait.

TABLE V-11

Computer Simulation of Human Judgments
For Five Essay Traits
(30 predictors, 256 cases)

A. Essay Traits	B. Hum.-Gp. Reliab.	C. Mult. R	D. Shrunk. Mult. R	E. Corr. (Atten.)
I. Ideas or Content	.75	.72	.68	.78
II. Organization	.75	.62	.55	.64
III. Style	.79	.73	.69	.77
IV. Mechanics	.85	.69	.64	.69
V. Creativity	.72	.71	.66	.78

NOTE:

Col. B represents the reliability of the human judgments of each trait, based upon the sum of eight independent ratings, August 1966.

Col. C represents the multiple-regression coefficients found in predicting the pooled human ratings with 30 independent proxes found in the essays by the computer program of PEG-IA.

Col. D presents these same coefficients, shrunken to eliminate capitalization on chance from the number of predictor variables (cf. McNemar, 1962, p. 184.)

Col. E presents these coefficients, both shrunken and corrected for the unreliability of the human groups (cf. McNemar, 1962, p. 153.)

The results here are not very surprising: mechanics shows the best agreement, and creativity the least, and this is in accordance both with intuition and other work on ratings. That is, English teachers are readier to agree on whether a word is misspelled, or an improper verb form used, than they are on whether a student's writing is original or shows imagination. We remember that spelling errors, inadequate as our list of misspellings is, nevertheless correlated -.30 with mechanics, while length of essay became a major contributor to creativity. It is clear, in any case, that human judges have a more difficult time with creativity than with other traits.

But what of the computer? Column C shows the raw multiple-R coefficients, predicting these criteria, unreliable as they are, from the prox measurements. Here we see that mechanics enjoys no advantage; to the contrary, it is more poorly evaluated by the computer than creativity is, and organization more poorly evaluated still. This relative standing, as we have seen, is contrary to the intuition of the humanist about what is easy, and what is hard, in the computer evaluation of prose.

Column D has made the reduction in MULTR which, as we have formerly discussed, is necessary to compensate for the capitalization on random error inevitable in multiple regression. These shrunken coefficients, then, have been found through statistical manipulation, or through lookup in Table IV-11(B), rather than through empirical cross-validation. Again, mechanics is not highest, and creativity not the lowest, of the shrunken correlations.

Column E exhibits a transformation of Column D, pumping up the correlation to compensate for the unreliability of the criterion scores. Column E, therefore, reflects the true population correlation which might be expected from the 30 proxes under the case of perfectly reliable judge ratings, and after eliminating the capitalization on random variation in the proxes. Thus the correlations in Column E are the best evidence to date about what success we theoretically would have in predicting the important qualitative dimensions of ideas, organization, style, mechanics, and creativity, using only computer-measured variables in the prose.

For all five traits, we have seen an ability to predict the "true" ratings with a rather surprising degree of accuracy.

Summary. This chapter has broken down the evaluation of essays into important dimensions, and has investigated strategies in predicting human judgments of these dimensions. New ratings were generated for 256 essays, with eight expert teachers, drawn from a sample of 32, independently grading each essay, on five traits commonly accepted as important. The judge intercorrelations, and trait differences, were shown. Then the chapter indicated how the proxes differentially contributed to the traits, so that spelling errors contributed to the evaluation of mechanics far more than they did to that of the other traits. Nevertheless, as one would expect from the halo effect demonstrated here by correlation and by analysis of variance, there was a great similarity in the lists of high contributors to the various traits. Some investigation was made of refining the criterion by gathering together similar judge viewpoints, and this possibility was recommended for further exploration. Finally, the overall ability of the system to predict the various traits was tested, and it was found that, contrary to what some might argue, such presumably lofty and subjective traits as creativity could be as effectively evaluated, using the present strategies, as well as the presumably more objective trait of mechanics. All in all, there did not appear to be any general area of essay evaluation which seemed, on any a priori grounds, beyond the possibility of automatic evaluation and analysis.

CHAPTER VI

PROBLEMS OF STATISTICAL IMPROVEMENT IN PREDICTION

The prior chapters have reported on work done over more than two years in analyzing essays mechanically. As has been seen, the success to date has been striking, although in a number of ways the reported strategies are surely less than optimum. One of the possibilities for improvement would appear to be in a more sophisticated strategy of statistical prediction. This chapter tells of explorations made into seeking some system other than the basic linear one of most multiple regression programs. Much of this chapter is based upon a report made by the authors (Paulus and Page, 1967) at an Annual Meeting of the American Psychological Association, in Washington, D.C.

The problem of linearity. A standard multiple regression program calculates an equation of the type shown as equation (1) in Figure VI-1. In this equation b_1 to b_{30} represent computer calculated weights for each of the proxes, x_1 to x_{30}. These weights are calculated in such a way so as to maximize the correlation between $\hat{Y}$ (the predicted score) and Y (the actual score or rating). We found this correlation to be over .65 (that is, on cross validation and after correction for attenuation). As we have seen from prior chapters, the method works. However, it does not work as well as it could, or perhaps should.

Of the many ways one might attempt to improve statistically upon the method, this chapter will report two, since both are applicable to a wide variety of multivariate predictive problems, not only to the grading of essays.

FIGURE VI-1

$$(1) \quad \hat{Y} = b_1x_1 + b_2x_2 + b_3x_3 + \ldots\ldots\ldots\, b_{30}x_{30} + c$$

$$(2) \quad \hat{Y} = (b_2x_2 + b_1)x_1 + c$$

$$(3) \quad \hat{Y} = b_1x_1 + b_2x_1x_2 + c$$

$$(4) \quad \hat{Y} = \sum_{j=1}^{30} b_jx_j + \sum_{j=1}^{30} \sum_{i=j+1}^{30} b_{ij}x_jx_i + c$$

$$(5) \quad \hat{Y} = b_1(x_1-\bar{x}_1) + b_2(x_2-\bar{x}_2) + b_3\,(x_1-\bar{x}_1)(x_2-\bar{x}_2) + c$$

$$(6) \quad \hat{Y} = b_1x_1 - \underline{b_1\bar{x}_1} + b_2x_2 - \underline{b_2\bar{x}_2} + b_3x_1x_2 - b_3x_1\bar{x}_2 - b_3x_2\bar{x}_1 + \underline{b_3\bar{x}_1\bar{x}_2} + c$$

$$(7) \quad \hat{Y} = \underline{b_1x_1} + b_2x_2 + b_3x_1x_2 - \underline{b_3x_1\bar{x}_2} - b_3x_2\bar{x}_1 + c^1$$

$$(8) \quad \hat{Y} = (b_1-b_3\bar{x}_2)x_1 + (b_2-b_3\bar{x}_1)x_2 + b_3x_1x_2 + c^1$$

Since both employ only existing data, the problems associated with the collection of further data are, therefore, avoided. It is our belief that some of the same problems will often haunt the workers with verbal data of the kind in this project, and therefore this discussion has relevance for other workers concerned with natural language strategies.

The first of the two approaches deals with the use of simple two-way interaction terms in an attempt to increase predictability. The second approach deals with the examination of the relationships between the various proxes and the criterion (the pooled ratings), with the thought of applying transformations to the proxes in an effort to increase the correlations between the proxes and the criterion. One of these uses interactions, then, and the second uses transformations.

Interactions. One way in which one might conceptualize an interaction term in a multiple regression equation is to think of variable weightings of predictor variables. We want the weights received by a given variable not to be a function of that variable's correlation with the criterion and the other independent variables alone, but also to be a function of the subject's score on some other variable. Equation (2) of Figure VI-1 will make this clear. Note that the weight received by x_1 in this simple equation is the quantity $(b_2x_2 + b_1)$; some function of the variable x_2 plus the constant b_1. Carrying out the indicated multiplication we obtain in equation (3) the simple cross-product of x_1 and x_2 which, along with the appropriate weight, represents the interaction of x_1 and x_2 on the criterion. Generalizing from this simple case, we can see that any number of cross-products (i.e., interactions) may be included in a multiple regression equation along with linear terms. Given our 30 proxes, then, we can look at 435 two-way interaction terms in addition to the 30 linear terms. This equation is of the form of equation (4) of Figure VI-1.

Before this equation can be calculated, however, two preliminary problems must be considered. The first problem is illustrated by equations (5) through (8) of that figure. Assume that we want to predict some criterion score Y from three independent variables: x_1, x_2, and the interaction of x_1 and x_2, x_1x_2. Equation number (5) gives the three predictor equation with independent variables expressed in deviation form. In equation number (6) the indicated multiplications have been carried out. Underlined terms are all constants and may, therefore, be absorbed in the constant term "C", as shown in equation number (7). After underlined terms have been combined in equation number (8) we find that the weight assigned to x_1 is the quantity $(b_1 - b_3\overline{X}_2)$ and the weight assigned to x_2 is $(b_2 - b_3\overline{X}_1)$. We can see, therefore, that the weights assigned to the linear terms in equation number (8) are distorted by the values of b_3, and $\overline{X}_1$ or $\overline{X}_2$. There are only two conditions under which this distortion is not present. First, when b_3 equals zero, and second, when the means of the linear terms are equal to zero. Since the interaction term is included precisely because the investigator does not believe that b_3 is equal to zero, only one alternative remains: to set $\overline{X}_1$ and $\overline{X}_2$ equal to zero. Further, and this appears to be contrary to all intuition, the correlation between the interaction terms is affected in a manner similar to the distortion of the weights assigned to the linear terms, unless the means of the linear terms are equal to zero. These distortions generally tend to inflate the correlations between the interaction terms and the criterion. Interaction terms, therefore, generally appear to be more valid than they really would be if the means of the linear terms had been adjusted to zero. If higher order interactions are considered, then the means of the lower order interaction terms must be adjusted to zero before the higher order interaction terms are calculated. Interestingly, the multiple correlation coefficient is not affected by these distortions; the crucial effect these distortions have is in the researcher's interpretation of his results. Hence, as the first step in working with interaction terms in the predictive context outline above, all linear variables must be standardized to a mean of zero. Another way of phrasing this dictum might well be: "No multiplication without standardization".

The second general problem which must be resolved, prior to the calculation of equation (4) mentioned before, is the selection of useful interaction terms. "Useful" is used here in the sense of an interaction term's ability to increase the multiple correlation. As mentioned before, we have, given 30 proxes, 435 possible interaction terms which could be included in the equation. It seems clear that not all of these interaction terms can be efficiently used in our predictive context. The reason for this is that, in cases where the predictors vastly outnumber the number of subjects, the loss in validity on cross-validation of the linear composite of terms becomes very, very great. A method of selecting useful predictors from all of the possible predictors needed, therefore, to be developed. A standard method usually employed in situations of this type is step-wise multiple regression. However, all step-wise multiple regression computer programs which we were able to find and to examine, required that at least one variable-by-variable matrix be stored in core memory of the computer. Given the amount of data available here, this would require a minimum of 200,000 core locations, too many for the computers currently available. As an alternative to the step-wise multiple regression procedure, the following method was employed.

A simple correlation coefficient is calculated between each of the independent variables and the criterion. The absolute values of these correlations are rank-ordered. The largest correlation is selected and the criterion is predicted from that variable which yielded the correlation. This is done for each subject. A new variable is then created which is the difference between the predicted criterion score and the observed criterion score. This new variable has the property of being uncorrelated with the independent variable which was just used. In other words, we now have a variable which does not correlate with the variable that was selected, nor with those portions of the other independent variables which correlate with the variable that was selected. In effect, a series of partial correlations are calculated: the first one being a zero order correlation, the next one a first order correlation, etc. At each step, the original criterion is replaced by the residual, the new variable, and the process is repeated until the residual correlates no longer with any of the remaining independent variables at some reasonable level, say, .05. This method provides for a rank-ordering of predictors. But the method has two weaknesses. First, the method is not as powerful as its converse. (Ease in programming, however, made the present method more desirable at this time.) Second, the method does not really allow for the selection of suppressor variables. This was unfortunate, and the investigators still seek a solution.

An additional problem inherent in all predictor selection techniques is that of cross validation. The validity of a multiple regression equation will, of course, almost always be highest for the sample in which the equation was constructed, and lower in other samples or in the population. As we have pointed out, formulas for estimating the cross-validities of sample multiple regression equations, such as the Wherry formula or the Lord-Nicholson formula, do not apply with complete rigor to situations where predictors were selectively chosen from a large number of predictors. Therefore, cross-validation estimates had best be established empirically. Thus, in this research, the previously described procedures were applied to a random sample of two thirds of our essays; the remaining third was used as a cross-validation sample.

Table VI-1 presents the data obtained when this cross-validation was applied. Note that only nine interactions and linear terms were selected before the correlation between the residual criterion and any of the predictors failed to exceed .05. These nine variables were entered into a standard multiple regression equation in order to obtain weightings and a measure of their combined predictive power. The obtained results are reported in Table VI-2. As expected, the multiple correlation is somewhat higher than the one obtained when the 30 linear terms were used. This increase, however, can't be evaluated until the equation is cross-validated and the amount of shrinkage has been discovered. Therefore, the obtained equation was applied to the remaining third of the sample, and the predicted scores were correlated with the observed scores. The correlation coefficient which was obtained was .63. You will note that this coefficient is approximately the same as the shrunken coefficient which was obtained by using the 30 linear terms - the proxes alone. This seems to indicate that we can predict the grade an essay receives as well from nine variables as we could from the original 30. Making use of interaction terms, therefore, does not allow us to predict any better, but rather to predict just as well using far fewer variables. The lack of increase in predictability is puzzling and may perhaps be attributed to the relative unstability of the criterion. If the criterion had been more reliable, this method would surely have yielded better results, for the reasons already explained in an earlier chapter.

TABLE VI-1

RANK-ORDERING OF PREDICTOR VARIABLES

Step	Variable	Correlation with Residual Criterion
1	Standard Deviation of Word Length	.52
2	Number of Commas	.23
3	Length of Essay in Words	-.15
4	Interaction of # of Periods and # of Subject-Verb Openings	-.15
5	Interaction of # of Periods and # of Declar. Sentences Type "A"	.12
6	# of Dashes	-.11
7	# of Words on Dale List	-.10
8	Interaction of # of Periods and # of Declar. Sentences Type "B"	.07
9	# of Connective Words	.05

TABLE VI-2

MULTIPLE REGRESSION EQUATION FOR LINEAR TERMS AND INTERACTIONS

Variables	B-Weight	Stan. Error	t	P
Linear Terms				
S.D. of Word Length	.0278	.0096	2.8958	$p < .01$
Number of Commas	.0050	.0014	3.5714	$p < .01$
Length of Essay in Wds.	.0023	.0006	3.8333	$p < .01$
Number of Dashes	.0223	.0097	2.2990	$p < .05$
No. of Connectv. Wds.	.0050	.0014	3.5714	$p < .01$
No. of Wds. on Dale Lst.	-.0191	.0092	2.0761	$p < .05$
Interaction Terms				
# of Periods X Subject Verb Openings	-.0014	.0008	1.7500	$p < .10$
# of Periods X # of Decla. Sents. Type "a"	.0011	.0005	2.2000	$p < .05$
# of Periods X # of Decla. Sents. Type "b"	.0203	.0099	2.0505	$p < .05$

Multiple Correlation = .74

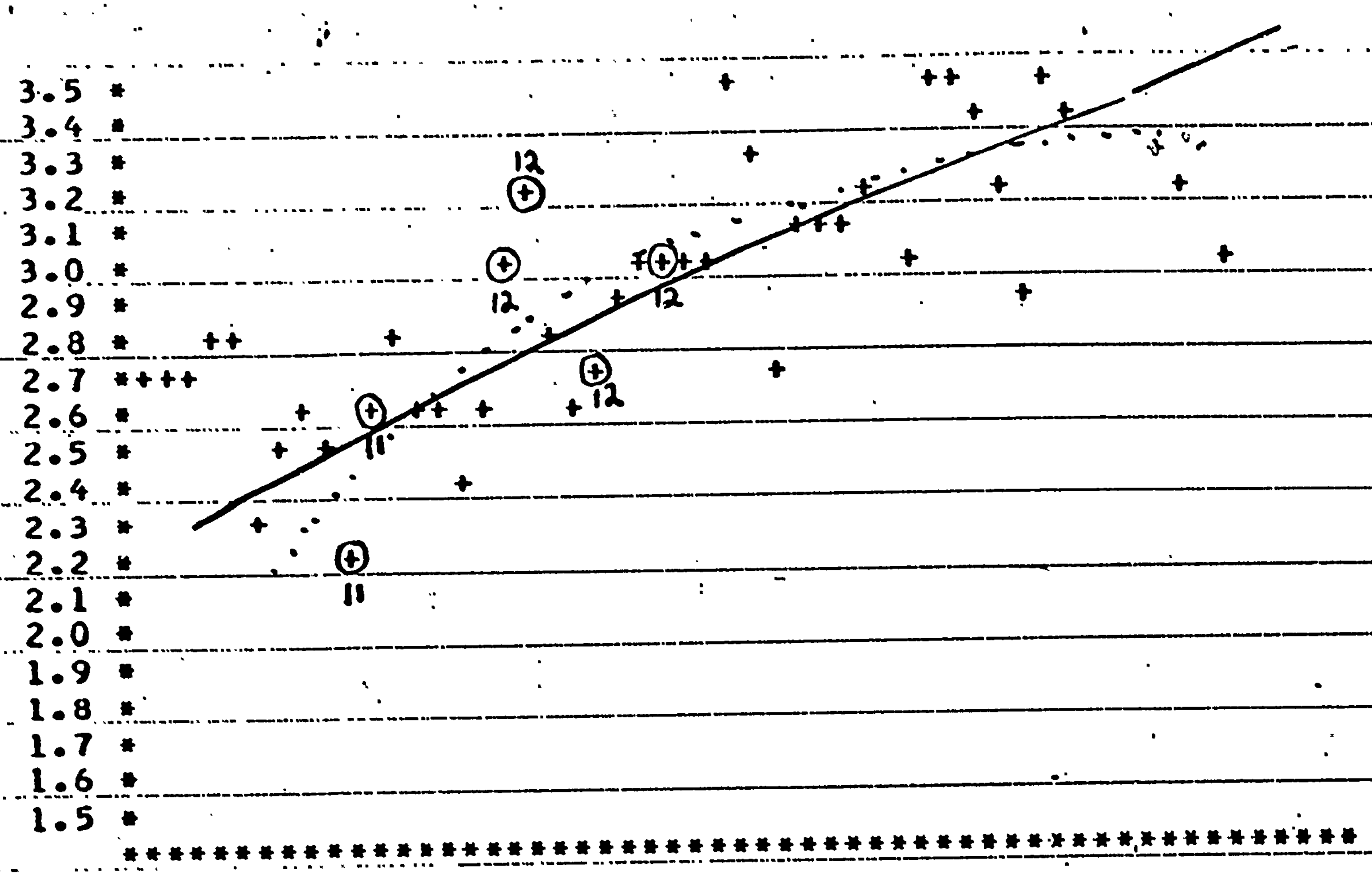
Stan. Error of Y = .471

Transformations. The first step in dealing with the relationships between each of the proxes and the criterion was to develop a short computer routine which would graph the relationship between each of the proxes and the criterion. Sample graphs for five variables are shown in Figures VI-2 to VI-6. It was hoped that by examining such graphs insights could be obtained which would aid one in selecting transformations to be applied to the proxes so as to yield higher correlations with the criterion.

On these graphs, the x or horizontal axis represents the independent variable, the prox. The y or vertical axis represents the ratings which an essay received. A rating of 1 is the lowest rating an essay could receive; a rating of 5 was the highest.

Each of the graphs was carefully examined in an effort to determine if any reasonable curve might explain the data better than a straight line. It appeared that for several of the graphs this might well be the case. Examine, for example, the graph for variable number 8 (Figure VI-2). The curve indicated by the dotted line may well fit the data better than the straight line. Both have been indicated on that graph. In order to sequentially apply transformations to the proxes, the following techniques were employed. A FORTRAN II program was written for the IBM 1620 computer (chosen locally for its auxiliary equipment and accessability) which allows a researcher to apply real-time transformations to the data. The program calculates means and standard deviations of both prox and criterion, and the correlation coefficient between the two. Next, the relationship between the two variables is plotted via a IBM 1627 plotter. These plots are similar to the ones in Figures VI-2 to VI-6, except that the points are connected and that a complete plotting grid is supplied. After examining the plot, the researcher can apply to the data one of

VARIABLE NUMBER 8 SCALE NUMBER 1

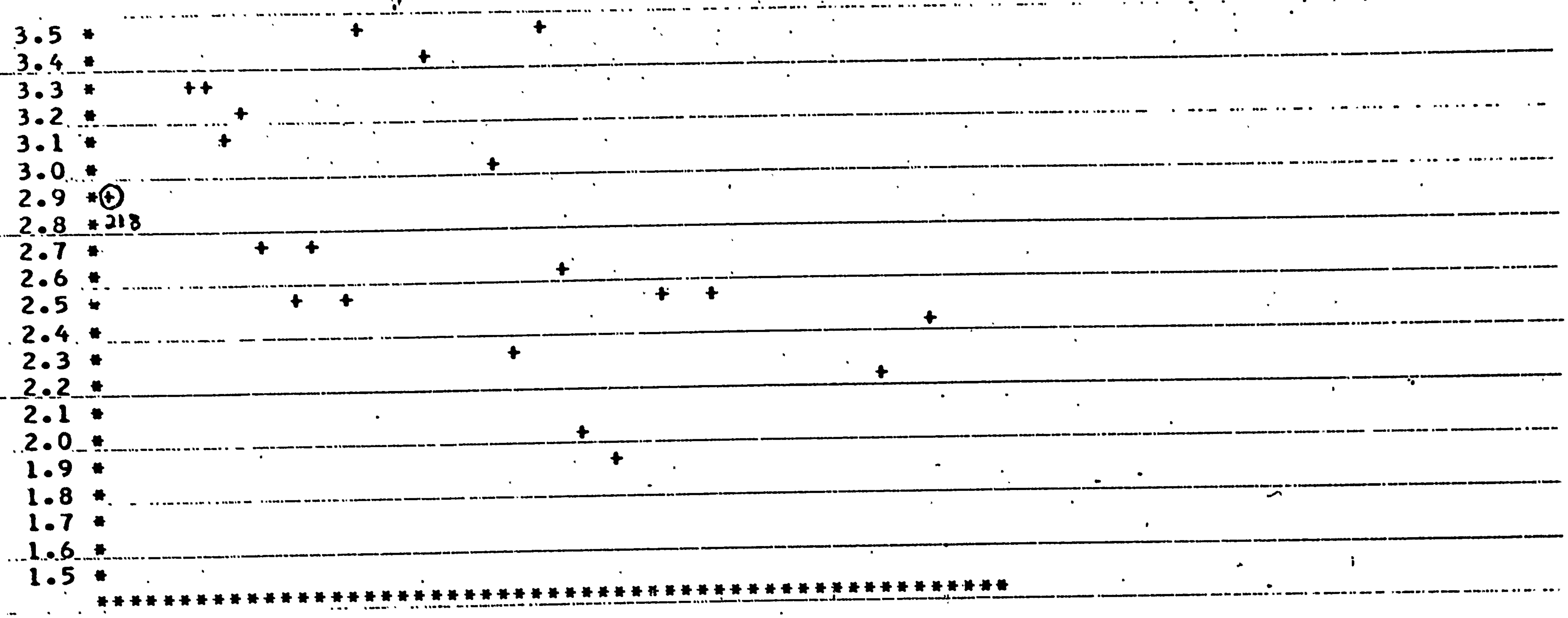


Number of Commas r = .34

FIGURE VI-2

Note: For this and subsequent graphs, the horizontal axis has been divided into 50 equal intervals. Points representing a substantial number of cases (usually 10) have been circled and the number of cases has been indicated.

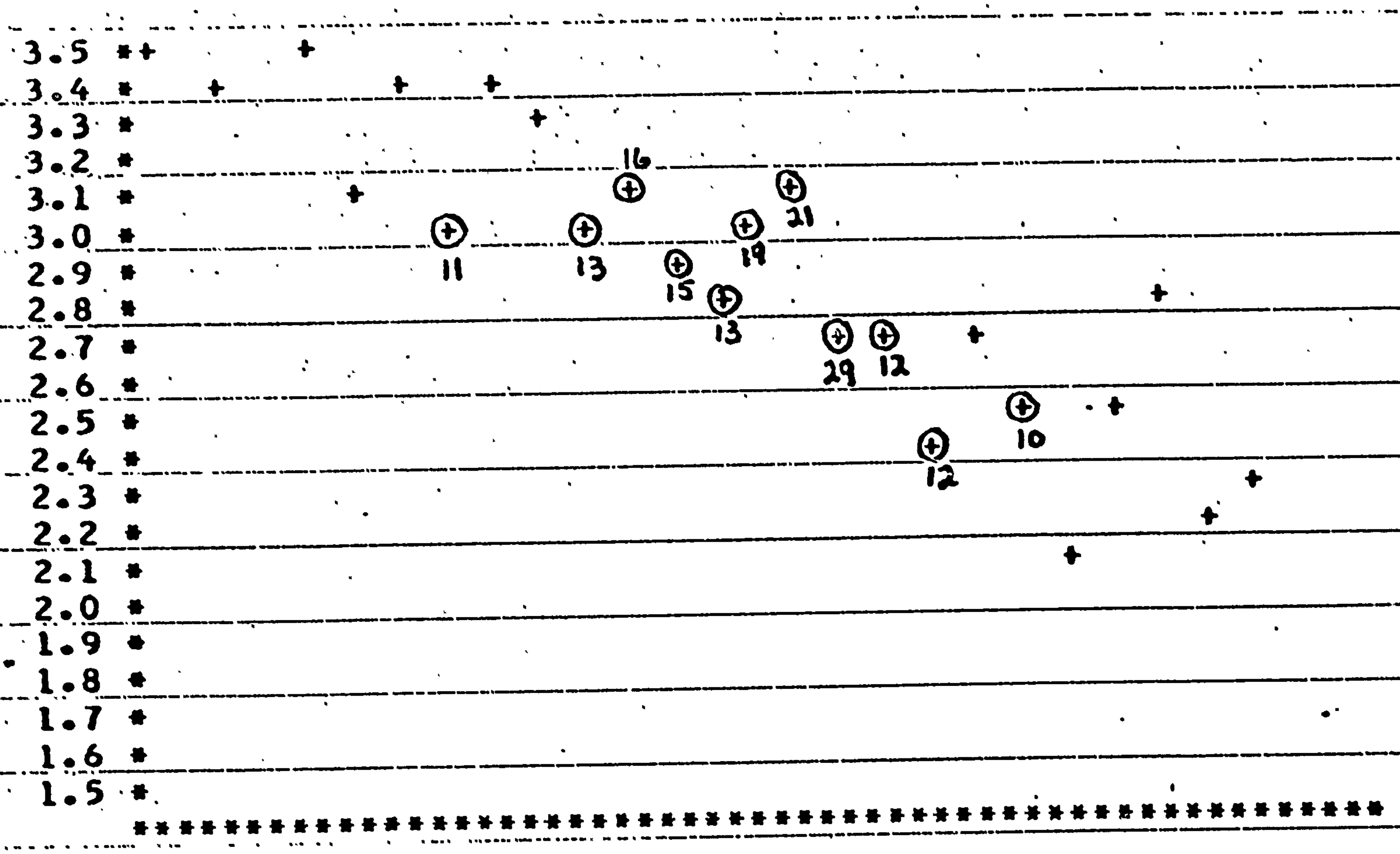
VARIABLE NUMBER 15 SCALE NUMBER 1



Number of Exclamation Marks r = -.10

FIGURE VI-3

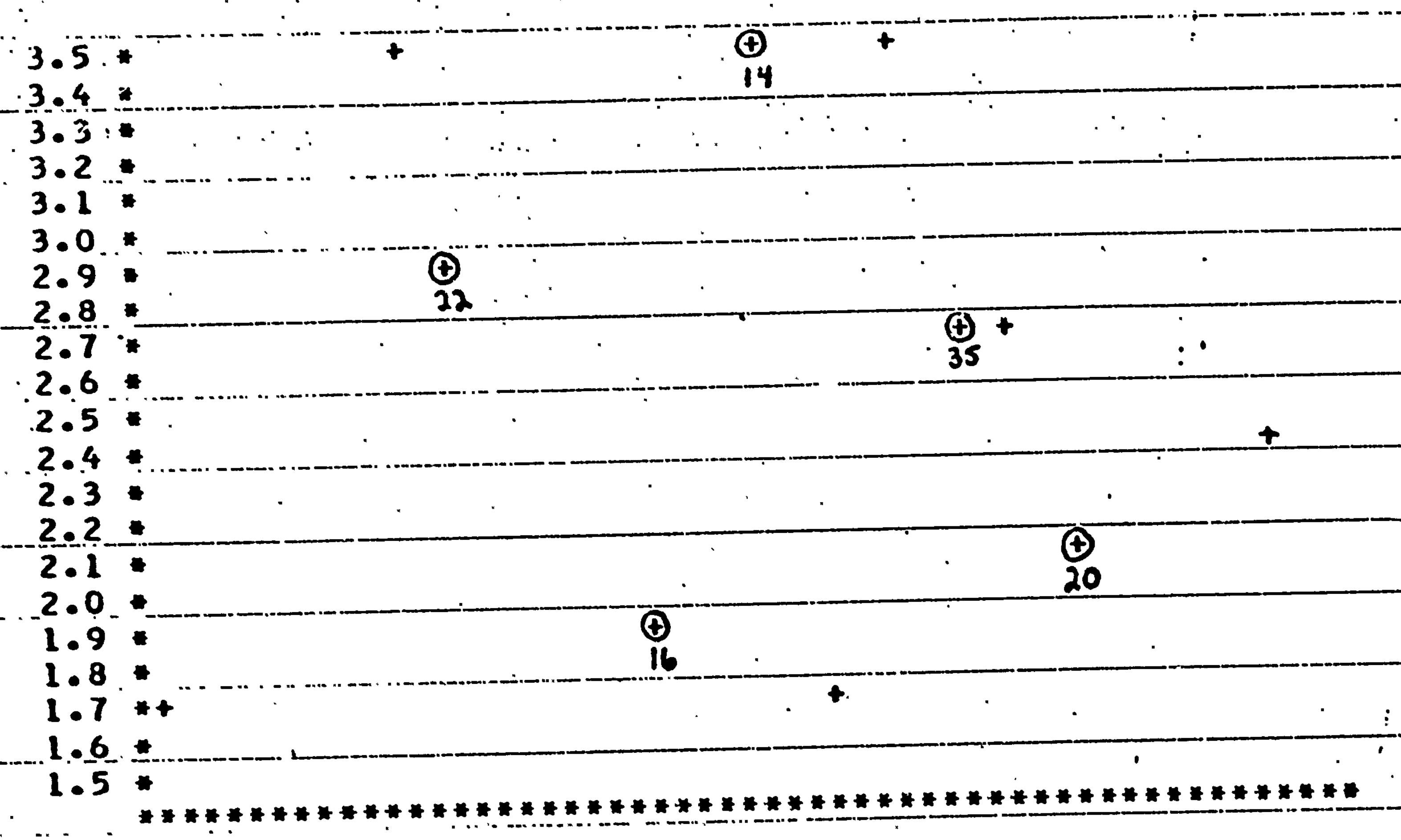
VARIABLE NUMBER 22 SCALE NUMBER 1



Number of Common Words on Dale List r = -.48

FIGURE VI-4

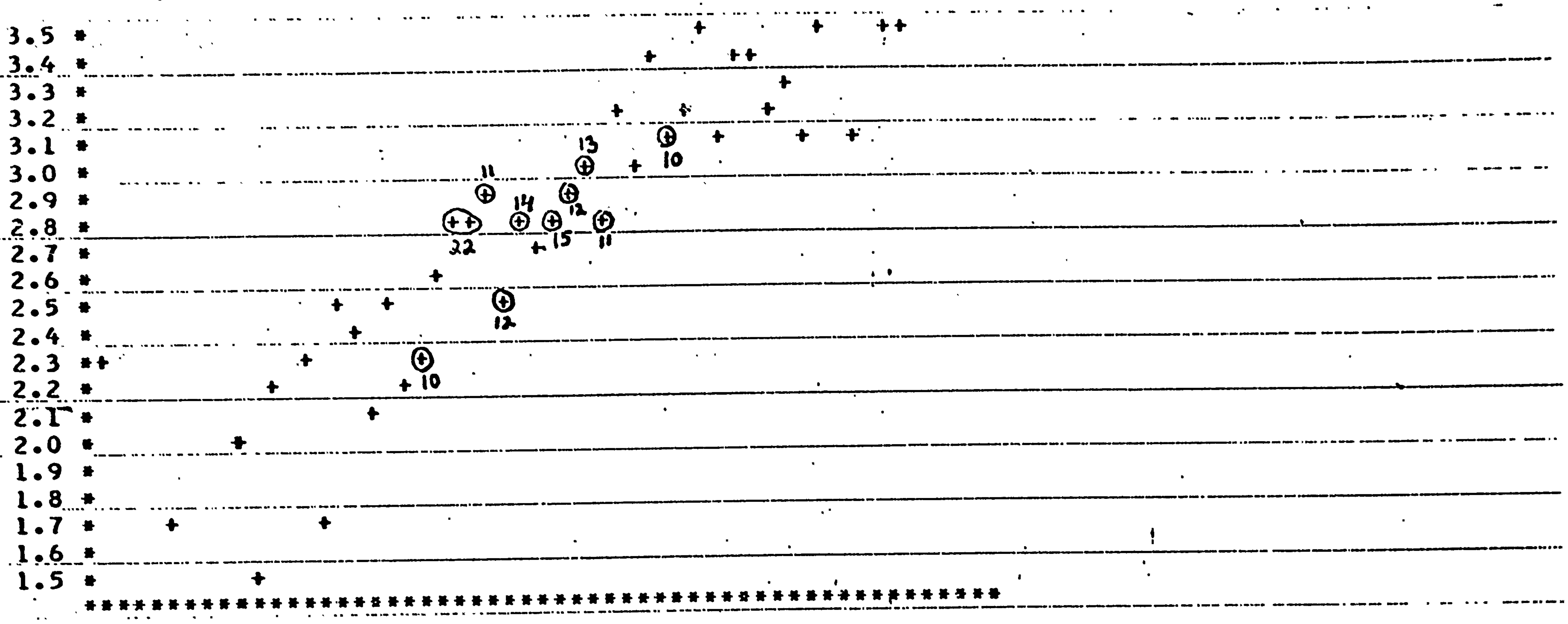
VARIABLE NUMBER 23 SCALE NUMBER 1



Number of Sentences with End Punct. Present r = .05

FIGURE VI-5

VARIABLE NUMBER 29 SCALE NUMBER 1



Standard Deviation of Word Length——r = -.52

FIGURE VI-6

14 transformations (or any combination of these transformations). The entire process is then repeated until the researcher decides to stop. The current values of the variables are then punched out on cards. This process is illustrated by Figure VI-7.

As the number of cases increases, this process becomes painfully slow on the 1620. (Compilation alone took about 20 minutes, and for 200 cases some transformations required as much as 15 minutes.) As a result, Paulus converted this program to run on our time-sharing teletype console, which is connected with an IBM 7094 at the Massachusetts Institute of Technology. Again, it appeared that the relative instability of the criterion limited the usefulness of this approach.

Discussion. To date, then, the Project has examined some methodological problems dealing with nonlinearity in predicting grades on essays. So far, however, we have not been able to substantially increase predictability by these methods, beyond that obtained under a naive linear assumption. It is our feeling that this may be due to the instability of the ratings of the essays, and, of course, the lack of more sophisticated proxes. Given the loose ratings used so far, it seems relatively unimportant what combinations of proxes are used, what transformations are applied to some of them, or what interactions are considered. The multiple correlation, after cross validation, appears to have stabilized at about .65.

There are at least two general ways in which such work may proceed in the future. The first is to recognize that there are differences among raters, and to attempt to empirically establish groups of raters, then to attempt to describe the characteristics of these groups. Some steps in this direction have been reported in the prior chapter. Then multiple regression equations, employing the previously

FIGURE VI-7

SAMPLE OUTPUT FROM TRANSFORMATION PROGRAM

```
CURVE FITTING PROGRAM                        PROBLEM NUMBER 076

THERE ARE 50 OBSERVATIONS

PLEASE CHECK SENSE SWITCH SETTINGS AND PRESS START WHEN READY

MEANS AND STANDARD DEVIATIONS

MEAN OF X =        1.9800

MEAN OF Y =        2.8890

S.D. OF X =         .9637

S.D. OF Y =         .5703

THE CORRELATION COEFFICIENT BETWEEN X AND Y IS       - .0720

IF STOP WRITE 9, ELSE 5

5

I AM READY TO ACCEPT ROUTINE NUMBER AND PARAMETER

13    2

MEANS AND STANDARD DEVIATIONS

MEAN OF X =        6.4369

MEAN OF Y =        2.8890

S.D. OF X =        2.0449

S.D. OF Y =         .5703

THE CORRELATION COEFFICIENT BETWEEN X AND Y IS         .1621

IF STOP WRITE 9, ELSE 5

5

I AM READY TO ACCEPT ROUTINE NUMBER AND PARAMETER

13    2

MEAN OF X =       52.6452

MEAN OF Y =         2.8890

S.D. OF X =       10.8134

S.D. OF Y =         .5703
```

FIGURE VI-7 (Cont.)

THE CORRELATION COEFFICIENT BETWEEN X AND Y IS .1423

IF STOP WRITE 9, ELSE 5

5

I AM READY TO ACCEPT ROUTINE NUMBER AND PARAMETER

6

MEANS AND STANDARD DEVIATIONS

MEAN OF X = 6.4369

MEAN OF Y = 2.8890

S.D. OF X = 2.0449

S.D. OF Y = .5703

THE CORRELATION COEFFICIENT BETWEEN X AND Y IS .1621

IF STOP WRITE 9, ELSE 5

9

END OF JOB

Note: All input is underlined.

discussed techniques, can be calculated for each homogeneous group of raters. We have recently completed a factor analysis of the 32 raters who rated our essays. However, since not all essays were rated by the same raters, we find that our data matrix contains more missing data than existing data. So we suspect that at least some of the factors which we empirically isolated are missing data and/or content factors.

A second general approach deals with differentially weighting raters before composite scores are calculated. This approach requires some judgment about the relative validity of each rater. Since we have no essays which have been rated by all of the raters, this poses some problems. One approach seems promising. This involves factor analysing the raters and using their factor scores (or some function of them) on the first principal component as weights.

Summary. In general, the investigators feel that workers with verbal data should be pleased but not contented with the present state of the art, and with the results obtained from using linear regression analyses. And they should continue linear analysis for the time being. But they should take care, whenever in doubt, to cross-validate the results. Further statistical optimization will probably be eventually profitable, when larger changes have been made in other aspects of the work.

CHAPTER VII

PHRASE LOOKUP AND ITS APPLICATIONS

The work described in the chapters up to this point has been limited by the computer program which has been used, and which has been shown in Appendix A. While this program, called PEG, is modular, mnemonic and flexible, it lacks any real convenience in looking up phrases. The present chapter describes a phrase look-up algorithm to accompany the program for essay analysis, and describes some studies done with the algorithm.

The phrase look-up procedure. The phrase look-up algorithm for this project was designed primarily by Donald R. Marcotte, and formed part of his M.A. thesis (Marcotte, 1966). Much of the present description is from his thesis or from the related report by Marcotte, Page, and Daigon (1967).

In one sense, of course, phrase lookup requires no special program. It is easy to insert in a FORTRAN program a conditional transfer of the form:

```
IF (WORD(I).EQ. X.AND.WORD(I+1).EQ.Y) GO TO ...
```

Here we have tested whether two words in a sequence of text words matched two words from some phrase. If the first text word in the sequence is not the same as X, then the test has failed, and the GO TO will not be executed. And if the first word is X, but the second text word is not the same as Y, then again the test has failed, and the GO TO will not be executed.

Such a test, however, lacks efficiency, and as a list of phrases of interest becomes large, would become very cumbersome to program, organize, and alter. What is desired is a procedure which permits search through a simply presented list of phrases, a list which may be regarded in the same way as the dictionaries in the main analysis program. And it is this need which the subroutine PHRASE was designed to fill. Appendix C has the source program listing for PHRASE.

In order to implement PHRASE, a skeleton copy of the PEG program was used to assemble the sentences of the essay being read, in the way already described. Also, the main program was used to read the array of first words of the phrases, and to read in the full phrase matrix.

One sentence is obtained from the essay being corrected. A phrase-within-quotation-marks (PWQM) counter, a PWQM indicator, and an adjusted word counter are set to zero. The PWQM counter is incremented every time a phrase is enclosed within quotation marks. The PWQM indicator provides a symbol, either 0 or 1, for punched-card output. The adjusted word counter eliminates unnecessary processing of words that have already been identified as part of a phrase. Since phrases of only two or more words are processed, the index indicating the number of words in the sentence is reduced by two, because the last word and end punctuation need not be processed.

DO LOOPS are set up which call the computer to cycle automatically until certain criteria have been met. The initial DO LOOP provides for the search of a sentence for a word that belongs to an array of first words of phrases. Prior to doing this, a test is conducted to determine if the index indicating the ordinal position of the word in the sentence is less than the value of the adjust word counter. If this index is less than the adjusted word counter no cycling occurs since the word being analyzed has already been processed, or it is the first word of the sentence. If the index is equal to or greater than the adjusted word counter, the word is processed. A provision is made to eliminate the processing of both parts of a natural-language word. This is necessary since a computer word (on the IBM 7040) consists of only six letters while many words in the English language contain more. Therefore, all natural-language words are represented by two computer words. This means that it is possible to identify the first part of an English word and also attempt to identify the second part of the same word. This possibility is eliminated by an appropriate test. The test is made by dividing the index for indicating the ordinal position of the natural-language word in the sentence. If the natural-language word has been processed previously, then no cycling occurs, and the next computer word is examined.

Because the computer cannot differentiate between natural-language words and punctuation marks, a test is conducted to determine whether the unit being analyzed is a punctuation mark. If this is so, no cycling occurs; but if the unit is not a punctuation mark, cycling does occur.

As was noted earlier, each natural-language word needs two computer words. Therefore the second DO LOOP requires two comparisons for each word provided to it. These two comparisons result in the identification of the particular phrase for which processing occurs.

After identifying the specific phrase, the adjusted word counter is incremented by two because two computer words have been processed. The index for the print-out array, RC, is set equal to two, and the two identified computer words are placed in the array, RC. The value of the row counter replaces the value of another row counter needed to process the phrase.

The third DO LOOP provides for the comparison of each natural-language word following the identified word with each natural-language word in the specific phrase. During each cycle a test is made to determine if the computer word of the sentence is the same as the computer word of the phrase. If it is, then the RC array index is incremented by one and the computer word is placed in the array, RC. If no comparison is made, then a test is made to determine if the symbol identifying the end of the phrase is present. If so, then indices for the identification of the presence of quotation marks are established. This is done in two steps: (1) by replacing the first index with the ordinal value of the natural-language word preceding the first word of the phrase in the sentence and (2) by replacing the second index with the ordinal position of the natural-language word succeeding the last word of the phrase in the sentence. The second index may have one of two values. This permits the identification of phrases that are not only enclosed within quotation marks but also have punctuation marks within the quotation marks. The first of the above alternatives is examined, and if the phrase is not enclosed solely within quotation marks then the second alternative is employed. If neither alternative is correct, then the phrase counter is incremented by one, and the computer word following the last natural-language word in the phrase in the sentence is cycled.

If in the test for an ending symbol, no comparison is made, then the index for the array, RC, is tested to determine if less than four natural-language words are in the array. The fourth word is not tested because phrases consisting of four words have no end symbol. If there are fewer than four words in the RC array, then the original row index counter is incremented and the next phrase is processed. This is done because several phrases begin with the same word, and it is necessary to examine all phrases having initial words in common. Continual incrementing of the row index counter occurs until all phrases beginning with the identified word in the sentence have been analyzed. This also means that one extra word will be analyzed, the initial word of the phrase following the phrases that have been examined, because the number of phrases beginning with the same first word are not constant. That is, it is not possible to determine when the series of phrases beginning with the same first word end. Therefore the added comparison is made.

Once the phrases are identified, it is necessary to record the information for "output." The output is punched on cards as well as printed on paper. There are two sets of punched card output: (1) the cards containing the identification number of the essay, the identification number of the phrase, the symbol indicating whether the phrase is enclosed within quotation marks or not, and the identified phrase, and (2) the cards containing the identification number of the essay, the total number of trite expressions used in the essay, and the total number of trite expressions enclosed within quotation marks. The printed output is an amalgamation of (1) and (2) above.

The final DO LOOP provides for the replacement of each word in the RC array by zero.

An application to cliches. Beyond constructing the described algorithm, the main purpose of Marcotte's study was to find how important cliches may be in the computer evaluation of student essays. Surely, according to English texts, such patterns of writing would be presumed to handicap an essay's evaluation, and might be expected to correlate negatively with human judgments.

Background on clichés. A cliché has been defined by Partridge (1962) as "...an outworn commonplace; a phrase, or short sentence, that has become so hackneyed that careful speakers and scrupulous writers shrink from it because they feel that its use is an insult to the intelligence of their audience or public." The searching of essays for clichés is a tedious if not impractical task. Certain clichés such as "each and every" and "null and void" seem to blend into a sentence so that they are not easily seen on the first reading. Second and third readings are often necessary to identify the cliché or clichés in the essay. The task, therefore, of spotting clichés seems insurmountable when there are several hundred essays to be examined, particularly when the essay has to be graded for other factors such as creativity, mechanics, style, organization, and ideas or content. The time required to make just one very detailed reading and commentary, a minimum of fifteen minutes (Daigon, 1966), is considerable, but when two or three readings are required the time multiplies greatly. Because clichés are clearly defined word groups, a computer search strategy is very efficient. Clichés can be stored in the computer and exact comparisons made.

LaBrant (1949) has discussed the difficulty of being sure when a cliché is hackneyed to the person using it, and Fowler (1965) has pointed out that every cliché seems fresh and novel at some time to the user. And Guth (1964) has warned against the "overzealous avoidance" of phrases which might seem trite, saying the there is a "not always clearly distinct borderline between the hackneyed and the idiomatic" (p. 194).

Partridge (1962), however, has approached the problem more systematically by providing a rather extensive list of clichés in dictionary form. He categorizes each cliché into one of four groups:

1. Idioms that have become clichés.
2. Other hackneyed phrases.
   Groups (1) and (2) form at least four-fifths of the aggregate.
3. Stock phrases and familiar quotations from foreign languages.
4. Quotations from English literature.

Other noteworthy aspects of Partridge's dictionary are definitive information and specific examples for each group of clichés, and the annotation of some clichés to indicate that these are considered particularly hackneyed or objectionable. Furthermore, Warriner (1951) and Griffith (1957), pp. 263-4) have supplied clichés not on Partridge's lists, and still others have been supplied by personal advance of Dr. Arthur Daigon. Three hundred clichés were included in this portion of the study, divided into two groups of 150 each: (1) clichés considered by Partridge to be "particularly offensive," and (2) others which were presumably not so odious. These lists may be found in Marcotte (1966, App. C), and will not be presented here.

Of the 256 essays examined, only 58 contained any occurrences of the cliché phrases, and there were only 74 occurrences all together. The number of different clichés used is only 24, and these are listed, together with their frequency of occurrence, in Table VII-1. An examination of these shows a rather large loading on two phrases: "finer things" and "in my opinion". When it is remembered that this particular essay was on whether, in a student's opinion, the best things in life are really free, it is understandable why these should occur so often. And these two phrases are seen to be pretty meaningless for any general conclusions. Of Partridge's "particularly offensive" phrases only eight were found, for a total frequency of only 13. In general, the clichés actually found do not seem necessarily very handicapping.

TABLE VII-1

TRITE PHRASES FOUND IN HIGH SCHOOL ESSAYS

Cliché	Frequency
all in all	3
by the same token	2
common understanding	1
each and every	1
finer things	14
first and foremost	1
helping hand	3
high and dry	1
in my opinion	19
in the long run	7
let's face it	1
matter of fact	1
more or less	3
really and truly	1
reigns supreme	1
root of all evil	2
step by step	1
survival of the fittest	2
this day and age	5
through and through	1
through thick and thin	1
to say the least	1
wishful thinking	1
work and no play	1

This intuitive feeling is borne out by Marcotte's statistical comparisons between those essays containing clichés and those not containing them. The mean ratings of the two groups (one with 58 essays, the other with 198 essays) were compared using a random-sample t-test, one-tailed because of the natural assumption that the non-cliché essays would be presumedly superior. No such evidence was found. To the contrary, for one trait (Ideas) the difference between group ratings was even in the wrong direction, and happened to account for the largest t-ratio found (1.37). And none of the t-tests approached significance.

We may infer that these particular lists of clichés, which are apparently as authoritative as any, do not aid in predicting whether an essay will be judged to have superior ideas, organization, style, mechanics, or creativity. Often findings of "no significant differences" are depreciated as inconclusive, or uninteresting to science. Here, however, where the data are drawn from a naturalistic essay situation and evaluated by realistic judges, such null-hypothesis findings seem to have great relevance. The avoidance of hackneyed phrases is often a subject of teaching in courses in composition, and this study casts a considerable shadow over the importance of the topic, at least in the secondary grades here sampled.

A search for psychological characteristics. Another application of the phrase look-up algorithm was in a study of what might be called quasi-psychological characteristics of prose (Hiller, Page, and Marcotte, 1967). This study was a combination of the strategies and methods used in this overall project, together with some of the subjective list-generation character of the General Inquirer (see Stone et al, 1966).

Traits were postulated which Hiller called "opinionation," "vagueness," and "specificity-distinctions", and for which he subjectively generated some phrase dictionaries, for intended use with the PHRASE subroutine already described. For the trait of opinionation, phrases were listed such as "I feel," "I think," "in my opinion," "who can doubt," etc., and the list included such apparent indicators of certainty as "all," "always," "beyond a doubt," etc., since opinionation and such certitude were believed to have something in common. All told, 130 items were included.

The other traits were similarly generated from an intuitive basis, supported by general admonitions in Strunk and White (1965). "Vagueness" was believed by Hiller to be indicated by such qualifiers as "probably," "usually," "a matter of opinion," "generally,: etc. This category of vagueness contained 60 items. And "specificity-distinctions" was believed to be indicated by words implying a specific, or concrete, point of view, such as "analyze," "ambiguous," "exception," "distinction," "specifically," etc. This list contained 90 words or phrases.

These phrase lists, then, were looked up in the 256 essays, and their correlations were studied with the same five traits of essay quality. To eliminate the general factor of length, the frequency of occurrences of such phrases should properly be divided by the total number of words of an essay, just as was done with other proxes. The correlations of these new proxes with the five trins are shown in Table VII-2. All correlations are in the predicted direction, and a number of them are highly significant, given the large number of essays represented. At first glance, then, the findings of Table VII-2 seem to lend some support for a kind of construct validity of the three traits postulated.

TABLE VII-2

CORRELATION OF FIVE MAJOR TRINS
WITH "OPINIONATION," "VAGUENESS," AND "SPECIFICITY"
(N = 256)

Trins	"Opinion."	"Vague."	"Specif."
1. Ideas	-.17*	-.26*	.08
2. Organization	-.20*	-.15*	.15*
3. Style	-.16*	-.22*	.10
4. Mechanics	-.14	-.14	.10
5. Creativity	-.14	-.32*	.04
Means	9.1	15.0	2.0
St. Deviat.	7.7	6.3	2.0

*Significant at the .01 level, with a one-tailed test.

Note: All correlations are based upon prox proportions, but the means and s.d.'s are raw frequencies.

Unfortunately, further analysis leaves the question very much in doubt, and the sum of evidence seems somewhat more negative than positive. It can be remembered from earlier chapters of this report that some of the largest correlates obtained with most of the trins were those proxes based on vocabulary: Dale common word list, average word length, and standard deviation of word length. These three are all presumably correlated with an inferred "frequency-score" of a student: the more words he uses which are infrequent, the more favorable will be these various vocabulary measures, and the higher will be his probable ratings.

These traits of "opinionation," "vagueness," and "specificity" were generated by uncontrolled subjective procedures, which would of course have no built-in safeguards against correlations with these other important proxes. Even the examples given here suggest a bias along the frequency dimensions: "I," "my," "always," "all," "probably," "usually," strike one as fairly commonplace, whereas "analyze," "ambiguous," "distinction," etc. are drawn from a less frequent set of terms. This supposition is borne out by the evidence in Table VII-3.

Here it is evident that the presumed dimensions are well enough correlated with prior vocabulary proxes so that the new evidence of correlation with essay quality does not contribute substantially in any search for construct validity for the new lists. If the lists happen to strike a reader as persuasive, then the measures, individual though they are, can be said to possess some face validity. But apparently we still do not have any more compelling evidence for their being important measures in their own right. This is a problem that is common in content analysis work. The problem is shared by the "dictionaries" used in most of the _General Inquirer_ work, as we

TABLE VII-3

CORRELATIONS OF VOCABULARY MEASURES
WITH "OPINIONATION," "VAGUENESS," AND "SPECIFICITY"
(N = 256)

Vocabulary	"Opinion."	"Vague."	"Specif."
Dale list	.32	.16	-.14
Aver. Wd. Leng.	-.43	-.06	.24
St. Dev. Wd. Leng.	-.18	-.19	.19

Only proportions are used for the column proxes.

have already noted, and the General Inquirer was clearly one of the two models for this sub-study.

More important, from an essay-analysis viewpoint, is the fact that the multiple-R for predicting essay quality does not seem to be increased by these traits of "opinionation," "vagueness," and "specificity." At first they seemed to one worker to contribute some new variance, but to date no cross-validation has shown significant improvement in the prediction through use of these hypothesized variables. This has some meaning for the major future developments in essay analysis, as will be discussed in a later chapter.

Correlative conjunctions. Some other types of routines have been developed for sequences of words which may be separated by other words. One worker, Alice Trailor, was curious about the use of correlative conjunctions, such as either . . . or, neither . . . nor, etc. Her reasoning was that sentences utilizing phrasal, clausal, parenthetical, or transitional elements would be indicative of a more mature or sophisticated style. And devices which provide means for coordination or subordination, such as correlative conjunctions, might be expected to predict human essay evaluations.

To test this relationship Miss Trailor used a lexicon of 11 common correlative conjunctions, taking Pence (1947) as a guide. For the 256 "free" essays, the resulting frequencies of such correlative conjunctions are shown in Table VII-4. Obviously, certain items dominated the usage of the high school students concerned, especially either... or and if...then, which together accounted for more than half of the occurrences. And with the judged quality of essays, these tiny frequencies had correlations hovering around zero, with the highest for any trait being a (non-significant) -.11 with rated creativity.

TABLE VII-4

DISCOVERED FREQUENCIES
OF CORRELATIVE CONJUNCTIONS

Correlative	Frequency
either...or	57
neither...nor	17
both...and	26
not only...but also	7
not only...but	8
if...then	44
although...still	3
although...yet	0
though...still	0
though...yet	0
since...therefore	0
	162

This particular investigation, then, explored one small facet of language usage in high school essays. The hypothesis that correlative conjunctions might furnish additional clues to writing quality was not supported by the data, but an algorithm was developed to permit the searching for separated words and word clusters in the text.

Verb constructions. Another investigator was interested in whether type of verb syntax would help predict essay quality. Thomas F. Breen pointed out that many textbook writers for composition teaching inveigh against the use of the passive voice, and claim that the active voice is almost always to be preferred (Gleason, 1965). But one would believe that perfect tenses, since they differentiate time, would characterize better writing (Scott, 1960).

Breen therefore developed an algorithm which would identify and count uses of the perfect tenses and the passive voice. His strategy was to locate the auxiliary verbs (forms of "have" or "be") and then look for a past participle (the algorithm searched for an -ed ending, or for membership on a list of 213 irregular past participles). Two general exceptions were noted: If a form of "be" were followed by a relative pronoun, then by a past participle, it was not counted as a passive verb. (Example: "There were many people (who) sent gifts.") Similarly, if a form of "have" were followed by the word "to," then by a past participle, it was not counted as a perfect form. (Example: "Someday you will have (to) come here.")

With the algorithm so developed, he found 367 occurrences of the perfect tenses, and 1323 occurrences of the passive voice. Of these latter, 135 were believed to be accompanied by a possible agent of the passive verb, a form generally regarded as worse than passive verbs not accompanying such explicit agents.

In general, Breen's hypotheses were not supported by the data. When the raw frequencies of such occurrences were correlated with the overall quality of essay, perfect tenses had a mere .03 relation. Passive voice occurrences had a correlation of .28 with essay grade, contrary to the prediction. And passive voice occurrences together with an agent had a correlation of .13 with essay grade. Unfortunately, the investigator did not control for essay length, which would expectably be correlated with these occurrences, and the discovered correlations are therefore harder to interpret than they might otherwise be. For example, if passive occurrences have a high correlation with essay length, and as we know essay length has a substantial correlation with essay quality, then the apparent correlation of passive occurrences with essay quality might be an illusion, and the meaningful correlation of the two variables might in fact be zero. And there are other possible third variables which would account for the apparent anomalies in the results. In any case, the project is turning toward a deeper syntactic analysis, as will be described in a later chapter.

Parenthetical Expressions. The final substudy described in this chapter was conducted again by Donald Marcotte. Parenthetical expressions are frequently used asides in writing. When properly employed, they are effective devices even though they do not contribute measurably to the over-all meaning of the sentence. The object of this section of the study is to determine whether the students used parenthetical expressions, and whether they used them judiciously. If so, then correlations between grades given on style and use of parenthetical expressions should be significant, and students using parenthetical expressions should receive higher grades than those not using them.

To test these hypotheses, we must first be able to identify a parenthetical expression. Fortunately, a parenthetical expression has two identifying features. The first identifying feature is its required punctuation. For example, Warriner and Griffith (1957, p. 580) state that "If he [the writer] wishes the reader to pause, to regard the expression as parenthetical, he sets it off; if not, he leaves it unpunctuated." Three types of punctuation marks are used in "setting off" the expression: commas, parentheses, and dashes.

The second identifying feature of the parenthetical expression is its placement in the sentence. According to Summey (1949, p. 60), there are three positions: "...(1) preliminaries, standing at the beginning of sentences or sentence members, (2) parenthetical groups in intermediate positions--commonly called parenthetical expressions with further qualification, and (3) tags or end parentheses."

With these two discernible cues, punctuation and position, and with a dictionary of parenthetical expressions, the computer can be programmed to identify these expressions in essays. The computer's dictionary consisted of 94 parenthetical expressions obtained from the textbook sources cited earlier, and from the opinionation-vagueness list already described.

Correlations and t-tests were used by Marcotte in the statistical analysis. First, correlations were computed to determine the relation between position of expression and grade given on style, and the relation between proportion of expressions used to number of sentences and grade given on style. Second, t-tests were used to determine if the group using parenthetical expressions received significantly higher grades on each of five traits (Ideas or Content, Organization, Style, Mechanics and Creativity) than the group not using parenthetical expressions.

Less than half (n=112) of the students used parenthetical expressions contained in the computer program dictionary. Of the 216 expressions found, 132 were used to introduce sentences, sixty-seven were used within sentences, and seventeen were used to end sentences. Also, commas accounted for the punctuation of 215 expressions; the remaining expression was set apart by parentheses.

Table VII-5 consists of a list of identified parenthetical expressions. Evidently, words like "also," "however," "no," "therefore," and the phrase "for example" are favorite items.

Table VII-6 shows the results of one-tail t-tests. All five comparisons were significant at the .01 level. However, the largest t-value was for style, as was expected. Apparently, the use of parenthetical expressions, proper use of course, has some bearing on the grades given on essays.

Table VII-7 shows the correlations between position in the sentence and style. Also shown are the correlations between proportion of number of expressions in the essay to number of sentences in the essay and style.

Except for the end position of the expression, all correlations are significant at either the .01 or the .05 level.

A summary of the Marcotte results, then, is as follows:

(a) One hundred twelve students used parenthetical expressions.

(b) Two hundred fifteen expressions were set-off by commas.

(c) One expression was set-off by parentheses.

(d) No dashes were used to punctuate the expressions.

TABLE VII-5

PARENTHETICAL EXPRESSIONS USED

Expression	Frequency	Beginning	Within	End
after all................	1	1	0	0
all in all..............	1	1	0	0
also.....................	15	11	0	4
at least.................	1	0	1	0
for example..............	17	14	2	1
for the most part.......	1	0	1	0
furthermore.............	1	1	0	0
generally...............	1	1	0	0
however..................	61	32	28	1
I am sure...............	2	0	1	1
I believe...............	2	2	0	0
I suppose...............	1	0	1	0
I think..................	1	0	1	0
if possible.............	1	0	1	0
in addition.............	1	1	0	0
in conclusion...........	3	3	0	0
in general..............	2	0	2	0
in my opinion...........	8	6	2	0
it seems................	1	0	1	0
likewise................	2	1	1	0
maybe....................	1	1	0	0
more or less............	1	0	1	0
moreover................	1	1	0	0
nevertheless............	3	3	0	0
no......................	15	10	0	5
obviously...............	1	1	0	0
of course...............	7	4	3	0
oh......................	5	5	0	0
on the other hand.......	8	3	5	0
ordinarily..............	1	1	0	0
perhaps.................	4	1	2	1
probably................	1	0	1	0
sometimes...............	4	4	0	0
still...................	1	1	0	0
that is.................	3	3	0	0
therefore...............	15	12	3	0
though..................	4	0	3	1
to be sure..............	2	0	2	0
too.....................	6	0	3	3
usually.................	2	2	0	0
well....................	7	6	1	0
why.....................	1	0	1	0

TABLE VII-6

MEAN GRADE DIFFERENCES
BETWEEN THE
PARENTHETICAL AND NON-PARENTHETICAL GROUPS

Traits	Difference between Means[a]	t	Probability
Ideas or Content	1.94	3.05	$< .01$
Organization	2.28	3.42	$< .01$
Style	2.43	4.02	$< .01$
Mechanics	2.71	3.57	$< .01$
Creativity	1.66	2.60	$< .01$

[a]Parenthetical minus non-parenthetical.

TABLE VII-7

CORRELATIONS BETWEEN

POSITIONS OF PARENTHETICAL EXPRESSIONS

AND STYLE

Position	$\bar{x}$	s	r	p[a]
Totals				
Beginning..................	0.516	0.863	0.23	<.01
Within.....................	0.262	0.599	0.21	<.01
End........................	0.066	0.293	0.00	n.s.
Total......................	0.844	1.154	0.28	<.01
Proportions				
Beginning/No. of Sent......	0.023	0.039	0.19	<.01
Within/No. of Sent.........	0.013	0.031	0.11	<.05
End/No. of Sent............	0.003	0.013	-0.02	n.s.
Total/No. of Sent..........	0.038	0.054	0.20	<.01

[a]One-tail test.

(e) Except for the end position, all correlations for position were significant.

(f) Students using parenthetical expressions received significantly higher grades on all traits than those students not using the expressions.

Again, it is wise to note some reservations about these findings. It is probable that all of the reported differences for parenthetical expressions might be affected somewhat by essay length. If an expression occurs in one essay but does not occur in other, it is likely that the one in which it occurs is a longer essay than the one in which it does not occur. And this relationship could have influenced the significance levels of Table VII-6. The second half of Table VII-7 attempts to provide for this influence, by controlling for the number of sentences in the essays. Nevertheless, this is not a wholly satisfying control, since sentences containing parenthetical expressions might be presumed longer than sentences not containing such expressions; and the factor of essay length might still be the major contributor to the observed relationship. Further multivariate study must be conducted to ascertain just how useful the discovery of these parenthetical expressions is going to be.

However, one portion of the present finding does not appear subject to this criticism of length, and is also very pleasing from an intuitive point of view. This is the contrast found, in the bottom half of Table VII-7, for the various positions of the parenthetical phrases. The correlation with quality of the proportion of beginning phrases is .19; of the within phrases is .11; and of the end phrases a (non-significant) -.02. This order coincides very nicely with the general view that end expressions are weak, dangling, and anti-climactic, and that middle expressions too often interrupt and divide the sentence syntax. This work on parenthetical expressions deserves further attention.

Summary. This chapter has made a significant extension in the facilities of essay analysis, by introducing a powerful and convenient phrase look-up subroutine, called PHRASE, written primarily by Marcotte, and tested with a number of sub-studies. One of these investigated the importance of cliches in predicting essay quality. It found that cliches were, first, surprisingly rare in occurrence in student papers and, second, quite inert in their apparent influence on ratings by expert human judges. A second study also used the PHRASE algorithm, with some subjectively constructed dictionaries, to investigate hypothesized traits of opinionation, vagueness, and specificity in the same student essays. Although found to correlate in predicted directions with essay quality, these three traits did not, apparently, contribute important unique predictive variance to the ratings. Other studies reported here investigated correlative conjunctions and verb construction in an effort to find predictors of essay quality. And a final study showed a positive relationship between writing quality and the use of parenthetical expressions, and their position in the sentence where used. In general, these uses of phrase procedures had varying degrees of success in the search for the sources of essay quality, but together they indicate the expanded utility of the essay analysis program.

CHAPTER VIII

ON-LINE ANALYSIS AND FEEDBACK

As we have seen from the earlier chapters, this project has repeatedly demonstrated that a computer can read a student's essay and return a numerical rating which indicates the quality of the essay on one of a number of traits. These ratings have been found to be as reliable as those assigned by trained human judges. Since the computer can return such ratings one might well ask, "What else can the computer return, given an essay as input? Can the computer make comments about a student's essay, and if so, on what can these comments be based?"

In an attempt to answer these and similar questions, a computer program was developed. This program, the interactive essay grader, instructs the computer to read a student's essay, to make a series of comments about the essay, and to allow the student to correct some errors which the computer found, all in conversation mode.

We should make it clear at the very beginning that this program is not to be taken as a model of expert pedagogy. The program requires much refinement before it can be used in a real school situation. The primary purpose of the program is to illustrate some of the things that can be done, and to reveal some of the problems which were encountered in its development. Most of this work has been carried out by Dieter Paulus, with some assistance by Michael J. Zieky, and was reported in much the present form to the American Psychological Association (Paulus, 1967).

Background. Since we are basically concerned with a simulation problem, simulating by computer the feedback an English teacher might provide for her students, a reasonable place to start would be in the examination of some of the comments an English teacher might make.

First, the teacher might look at the content of the essay to see whether the student has demonstrated an understanding of the required concepts and a knowledge of the required facts. At present, no attempt was made to program the computer to comment on the content of the student's essay. (There is work now beginning in this field.)

Second, a teacher may look at the general structure of the essay. Here the teacher might be concerned with the soundness of the inferences a student draws, whether or not the mode of expression used by the student is appropriate, where the essay lacks clarity, or where a point needs further support. Here again, no efforts were made in this program to allow the computer to deal with these areas.

A third aspect of an essay that a teacher may look at in a student's essay, and frequently this is the most important and most time consuming task in which an English teacher is involved in the teaching of elementary writing skills, is the judging of the appropriateness of the student's word usage, determining errors in declension, noting spelling errors, and so on. Comments relating to these areas are appropriately made if the writing of an essay is seen primarily as a drill exercise, and the student is asked to write many essays so that he may learn to avoid these errors. It is the comments a teacher writes relative to these types of errors that the present computer program attempts to simulate; for these comments are rather routine and take up much of the teacher's time and energy. If the computer can successfully take over this task, then it would be doing the teacher a tremendous service, as she could spend her time and energy in making comments of the first two types.

The type of feedback which this program attempts to simulate is of the prescriptive sort, comments that tell the student to avoid certain usages and suggests certain alternatives. If the student's essay deviates too greatly from the norm, then the computer indicates to the student where potential problems may lie and suggests corrective measures.

The program. The interactive essay grader is, with the exception of one short subroutine, written entirely in FORTRAN IV. The program was written on a remote teletype terminal, connected by telephone cabel to M.I.T.'s IBM 7094 computer. A student who wishes to use the program simply types a code word on the console and the program begins to execute. At the appropriate point in the execution of the program, the computer asks the student to write the essay in natural language. The only restriction imposed on the student are special punctuation marks. This is due to the limited character facility of FORTRAN IV. When a subject has completed the essay, he is instructed to type an asterisk. The computer then starts almost immediately to respond and to comment on the student's essay.

As a language, of course, FORTRAN IV is not particularly well suited for natural-language computing. Therefore, the program in its present form is relatively inefficient and lacks elegance. Nevertheless, the computer requires only about twelve seconds of machine time to evaluate and to begin comment on an essay. Printing speed is, of course, considerably slower.

For purposes of describing the program, it may be conveniently, though artifically, divided into five parts. These are (1) the grading routine; (2) the prescriptive comments; (3) comments based on actuarial characteristics of the essay; (4) the interactive spelling foutine; and

(5) the recalling and recording of data about the essay. Each of these will be discussed in turn.

The program calculates a numerical grade for an essay by using a weighted sum of scores on eight variables. These variables were selected by a step-wise multiple regression process from the original 30 proxes used in the larger project. The eight variables which are included yield a multiple correlation coefficient of approximately .60 when used to estimate expert human ratings. The program takes the numerical grade and selects an appropriate comment from a list of comments. If, for example, the grade is quite high, the computer writes, "I think that you did quite well. Keep up the good work!" On the other hand, a very low grade calls for the response, "I don't think that you did at all well. Are you taking this assignment seriously?" Intermediate grades call for other comments. (Incidentally, if a student tries to fool the computer and types nothing but nonsense, the computer responds, "Stop wasting my time! If you don't stop playing around I will report you to your teacher".) These comments are used instead of numerical scores because they are presumedly more meaningful to the student than, say, the number 2.8634. If teachers usually had time to write comments, they undoubtedly always would. The number or letter grade alone is primarily designed to save time.

The prescriptive comments are called by a binary search phrase look-up subroutine which search lists that have previously been entered into the computer's memory. Michael Zieky was primarily responsible for developing this portion of the program. The lists which can be searched by the computer may be of almost any length, limited only by the size of the computer. Since search time is not directly proportional to the length of the lists, these lists can grow to great lengths with only a trivial loss in computing time. For example, to search a list containing 16,000 words requires only one more comparison by the computer than to search a list of some 8,000 words. Hence the criticism as to why a particular word or phrase is not included in the list is quickly dispelled by simply including that word or phrase.

The general classes of words and phrases included in this list and on which the computer comments are as follows: (1) Taboo words, such as "aint" or "busted"; (2) misuses of case, such as "theirselves" or "to who"; (3) use of "of" for "have", "could of" or "should of", for example; (4) noun-verb disagreement, for example "both is" or "I are"; (5) misuse of homonyms, "their is", for example; (6) vulgar idioms such as "somewheres" or "that there"; and (7) double negatives such as "can't hardly" or "don't scarcely". As indicated before, it is only the researcher's knowledge and imagination that limits the classes and number of words or phrases to be included.

If the computer finds an improper usage it prints a message. For example, if the word "irregardless" is encountered by the computer, it responds, " 'Irregardless' is actually a double negative. If you examine the first and last syllables you will see why.", or if the student writes "should of" the computer responds, "When we speak quickly the word 'have' often sounds like 'of'. But it should never be written that way." If the student writes "busted" the computer responds, "Do you really think the past participle of 'break' is 'busted' or were you just being careless?"

Comments based upon actuarial characteristics of the essay are printed whenever some characteristic of the essay deviates too greatly from its normative use. For the time being, norms are based on a sample of 256 essays used in previous analyses and can readily be changed as the type of essay changes.

These comments are generally stated so as to indicate to the student that there may be a problem with given aspects of his essay. The computer might say, for example, "Your sentences seem long and complicated...." and ask the student a question, or suggest how the difficulty may be overcome. That is, the computer indicates that there may be a problem and suggests that the student check to see whether or not a problem really exists.

The interactive spelling routine again utilizes a binary search to determine which words are misspelled. The present list includes some 750 words. First the computer prints a list of the misspelled words that it found, then it gives the student an opportunity to correct them. The student is asked to spell a given word correctly, and if he does so, the computer responds "That is correct. Very good." If the student continues to spell the word incorrectly, the computer first suggests that the student try again, then, if it is again incorrectly spelled, that he look the word up in a dictionary. If the student again makes an error, the computer finally suggests that the student go and seek his teacher's help; then it goes on to the next word. The computer determines whether or not the word is spelled correctly by looking the word up in a list of correctly spelled words corresponding to those spelled incorrectly in the spelling list.

There are several problems inherent in this procedure. First, the list of approximately 750 misspellings seems to be quite inadequate. This judgement is based on the examination of a glossary of the words used in 256 essays written by high school students. It was discovered that only a fraction of the misspellings found in those essays appeared on the spelling list. But again, the list can be easily increased in length. Second, it is sometimes difficult to determine whether or not a word should be included in the spelling list at all, since some commonly misspelled words correctly spell some other words. For example, if the word 'M U S S E L' is included as a possible misspelling for the word 'M U S C L E', then if the student really intended to spell 'Mussel' it will be counted as a misspelled word. A further problem involves a student who misspells a word not in the anticipated manner, in a plural form for example. Then the computer will not recognize the word as a misspelling. A partial solution to these problems may lie in the inclusion of an extensive dictionary of correct spellings along with given rules for forming plurals, possessives, etc. which may be applied by the computer. This approach is currently under investigation by Francis Archambault.

The last part of the program deals with the recording of data about a student's essay for use in making comments on future essays by that same student and for reporting the student's progress, or lack thereof, to his teacher. These data are recorded on a disk and are always available to the computer. The computer can, therefore, look back to the student's previous performance and compare his present performance to that. For example, when commenting on a student's overall grade, the computer can add, "You did much better than last time. Very good!", or, if the student makes a greater number of grammatical or word usage errors, the computer may comment, "With respect to grammar and word usage you have done considerably worse this time than last time". Similar comments are made when the total number of spelling errors are reported. If a student makes the same spelling error in two consecutive essays, the computer comments, "By the way, you made this same error the last time that you wrote an essay for me. Please be more careful."

These data are also used to provide the teacher with feedback about a student's essays. By typing the appropriate code word and code number on the teletype console, the teacher may obtain a summary of the grades on all of the essays a student has written, as well as other summary information about various errors the student has made in each essay that he has written. If the teacher wishes, she can obtain complete copies of essays any student has written.

Figure VIII-1 will show some sample computer output. The computer's comments are designated by ordinary type. The student's essay and responses are in italics. The essay which the computer evaluated was written by Paulus to demonstrate some features of the program.

Summary. A computer program has been developed which provides limited feedback to both the student and to the teacher about a student's essay. Content and general stylistic considerations are not considered as yet, but the program now comments on some errors in grammar and word usage, as well as on certain actuarially derived variables. A great deal of flexibility is provided for increasing the number of errors which the computer can recognize and comment upon. The program is, as of now, untried in real classroom situations. But it does illustrate a number of interesting possibilities and indicates that this general area is a fruitful area for further research. The computer can provide relevant feedback to the student about his essays. This feedback can deal with a variety of topics and can take many different forms.

Future steps should eventually include the refinement of the spelling routine, the development of longer and more comprehensive lists of possible errors, refining and increasing the number of comments, and re-writing the entire program in a more suitable language, such as PL/1. Teachers should also have greater control over comments and when they are printed. Carl Helm's PROTRAN, or a similar language, appears to be admirably suited for such a purpose.

FIGURE VIII-1

(Note: The "student's" responses are in italics)

I AM AN IBM COMPUTER. WHAT IS YOUR NAME PLEASE.Q

JOHNNY H. DOE

DID YOU EVER WRITE AN ESSAY FOR ME BEFORE.Q

YES

HELLO, JOHNNY H. DOE, IT IS NICE TO TALK TO YOU AGAIN. PLEASE TYPE THE CODE NUMBER ASSIGNED TO YOU THE LAST TIME WE TALKED.

89

THANK YOU. WOULD YOU PLEASE READ THE DIRECTIONS ON THE FRONT OF THE MACHINE NOW, AND BEGIN WHEN READY.

THE NAME OF MY FAVORIT TELEVISION PROGRAM IS BATMAN. I WATCH IT EVERY CHANCE I GET EXEPT LAST WEEK WHEN OUR TV SET WAS BUSTED. THEN I COULDN'T SEE IT.

THE WEEK BEFORE THAT THEY HAD A REALLY GOOD SHOW ON. THE CATWOMAN STOLE SOME JEWELERY SHE LIKED. SHE HID THE STUFF SOMEWHERES. I BELEIVE THAT IT WAS IN HER HIDE-OUT.

BATMAN AND ROBIN AINT STUPID SO THEY SOON FOUND IT AND ALMOST CATCHED THE CATWOMAN. SHE GOT AWAY DURING THE FIGHT. THEY COULD HAVE FOUND HER LATER BUT THEY DIDN'T WANT TO GIVE THEIRSELVES AWAY.

THEIR IS OTHER NEAT CROOKS ON THE PROGRAM TOO. LIKE THE PENGUIN THE RIDDLER AND THE JOKER.

THEY ARE ALWAYS TRYING TO GET THE DYNAMIC DUO, BUT THEY NEVER SEEM TO BE ABLE TO.

***

7094**7094****7094****7094****7094****7094****7094****

HERE IS SOME INFORMATION ABOUT YOUR ESSAY.

7094**7094****7094****7094****7094****7094****7094****

I DON'T THINK THAT YOU DID VERY WELL.
TRY HARDER NEXT TIME AND PAY CLOSER ATTENTION TO WHAT I TELL YOU.
(YOU DIDN'T DO AS WELL AS YOU DID LAST TIME.)

Continued

(IF YOU DON'T AGREE WITH THIS, OR WITH ANY OF MY OTHER COMMENTS, CHECK WITH YOUR TEACHER.)

IN RESPECT TO GRAMMAR AND WORD USAGE, YOU HAVE DONE ABOUT AS WELL THIS TIME AS YOU DID LAST TIME YOU WROTE AN ESSAY FOR ME.

DO YOU REALLY THINK THE PAST PARTICIPLE OF "BREAK" IS "BUSTED", OR WERE YOU JUST CARELESS.Q

"SOMEWHERES" IS AN EXAMPLE OF POOR SPEECH HABITS SHOWING IN YOUR WRITING. THE WORD SHOULD BE PRONOUNCED AND SPELLED WITHOUT THE FINAL "S".

I NOTICE YOU USED THE WORD "AINT". MANY PEOPLE STILL OBJECT TO IT. SEE IF YOUR TEACHER ALLOWS THAT WORD BEFORE YOU USE IT AGAIN.

SOME VERBS ARE IRREGULAR BECAUSE THEY FORM THEIR TENSES IN A STRANGE WAY. "CATCH" IS SUCH A VERB, AND ITS PAST TENSE SHOULD BE "CAUGHT", NOT "CATCHED".

THE WORD "THEIRSELVES" SHOULD BE REPLACED BY "THEMSELVES". SEE IF "THEIRSELVES" IS IN YOUR DICTIONARY.

I SEE THAT YOU HAVE USED "THEIR" FOLLOWED BY A VERB. THE CORRECT FORM IS "THERE". WORDS THAT SOUND ALIKE BUT ARE SPELLED DIFFERENTLY ARE CALLED HOMONYMS. CAN YOU THINK OF ANY OTHERS.Q

YOUR ESSAY IS CONSIDERABLY SHORTER THAN THE ASSIGNMENT REQUIRED. HAVE YOU FULLY DEVELOPED THE TOPIC.Q

YOUR SENTENCES SEEM SHORT AND CHOPPY. COMBINE SEVERAL USING APPROPRIATE CONNECTIVES SUCH AS 'AND' OR 'OR', FOR EXAMPLE.

YOU USED 5 PARAGRAPHS.

WHY SO MANY.Q TRY TO REORGANIZE THE ESSAY SO THAT YOU DON'T USE AS MANY PARAGRAPHS. USUALLY, A PARAGRAPH SHOULD CONSIST OF SEVERAL SENTENCES, AND SHOULD PRESENT ONE COMPLETE UNIT OF THOUGHT.

I NOTICED THAT YOU DIDN'T USE VERY MANY COMMAS. DID YOU SEPARATE ITEMS IN A SERIES BY COMMAS, USE COMMAS IN COMPOUND SENTENCES, USE COMMAS AFTER INTRODUCTORY PHRASES.Q PERHAPS YOU HAD BETTER GO BACK TO YOUR ESSAY AND CHECK THOSE THINGS.

Continued

I AM NOT CERTAIN BUT I THINK THAT
YOU MADE   4 SPELLING ERROR(S).

THAT IS A GREATER NUMBER OF ERRORS THAN
YOU MADE IN YOUR LAST ESSAY. PLEASE BE MORE CAREFUL.X

HERE IS A LIST OF YOUR SPELLING ERRORS.

FAVORIT
EXEPT
JEWELERY
BELEIVE

(THERE MAY BE OTHER ERRORS THAT I DID NOT FIND.)

LET'S SEE IF WE CAN CORRECT THE ERRORS I FOUND.

LOOK AT THE FOLLOWING WORD.

FAVORIT

PLEASE TRY TO SPELL THE WORD CORRECTLY.

(TYPE ON THE NEXT LINE AND PUSH 'RETURN' ONE TIME.)

FAVORITE

THAT IS CORRECT. VERY GOOD.

LET'S GO ON TO THE NEXT WORD.

LOOK AT THE FOLLOWING WORD.

EXEPT

PLEASE TRY TO SPELL THE WORD CORRECTLY.

EXCEPT

THAT IS CORRECT. VERY GOOD.

LET'S GO ON TO THE NEXT WORD.

LOOK AT THE FOLLOWING WORD.

JEWELERY

PLEASE TRY TO SPELL THE WORD CORRECTLY.

JEWLERY

Continued

NO, THAT IS STILL NOT CORRECT.
WOULD YOU PLEASE TRY AGAIN.

JUWELERY

I AM SORRY. THAT IS NOT CORRECT.
WOULD YOU PLEASE LOOK THE WORD UP IN THE DICTIONARY AND TRY AGAIN.Q

JEWELRY

THAT IS CORRECT. VERY GOOD.

LET'S GO ON TO THE NEXT WORD.

LOOK AT THE FOLLOWING WORD.

BELEIVE

(BY THE WAY, YOU MADE THIS SAME ERROR THE LAST TIME THAT YOU WROTE AN ESSAY FOR ME. PLEASE BE MORE CAREFUL.X)

PLEASE TRY TO SPELL THE WORD CORRECTLY.

BELIEVE

THAT IS CORRECT. VERY GOOD.

WELL, I HOPE THAT TAKES CARE OF THE SPELLING ERRORS.
PLEASE TRY NOT TO MAKE THE SAME ONES AGAIN.

7094**7094****7094****7094****7094****7094****7094****

WELL, JOHNNY H. DOE, IT WAS NICE TO
TALK TO YOU AND TO READ YOUR ESSAY.

I HOPE THAT YOU WILL COME BACK SOON TO WRITE ANOTHER ONE. MEANWHILE, PLEASE THINK ABOUT WHAT I TOLD YOU.

GOOD BYE.

7094**7094****7094****7094****7094****7094****7094****
7094**7094****7094****7094****7094****7094****7094****

PLEASE NOTE YOUR NEW CODE NUMBER WHICH IS 90.
THANK YOU.

DO YOU WANT TO WRITE ANOTHER ESSAY NOW.Q
PLEASE ANSWER 'YES' OR 'NO'. (NO BLANKS.)

NO

EXIT CALLED.

Even though much work remains, and many problems are as yet unsolved, the interactive essay analyzer designed by Paulus seems to have opened the door to research in a relatively new aspect of computer assisted instruction, an aspect of computer assisted instruction that allows the computer to assume a greater role than that of a "mechanized scrambled book". The computer begins to understand what it is told by the student and is able to intelligently respond to him. Such on-line work should eventually become an important area of application.

CHAPTER IX

CONCLUSIONS AND IMPLICATIONS

Chapters I through VIII have discussed rationale, methods of empirical research, and various findings from the work to date. However, the investigators have recognized from the beginning the extreme newness of this study, and its vast potentialities for the future of educational measurement and instruction. Consequently, part of the original charge of this project was to scan the field constantly for new opportunities of research and practice. Some of the recognized opportunites will grow rather directly out of the work so far accomplished within the project, but others will stem from sythesis with other work in related fields. Therefore this chapter will perform three functions: (1) It will summarize the preceding chapters and the major line of work within this project. (2) It will discuss work in tangential fields, and the general status of the disciplinary interface most appropriate to the future of essay analysis. (3) It will point out some appropriate directions for future work within the field of educational measurement and instruction, future work which may be closely related to this project.

1. Summary of Work Completed

Rationale. The basic strategies of the computer analysis of essays have all grown out of an attempted simulation of human ratings. The fundamental approach has been to seek a goal of automatic analysis of stylistic qualities in essays, and the techniques have been generally actuarial. That is, we have looked for a simulation of human, expert judgment of intrinsic qualities (trins), through an exploration of correlated, or approximate variables (proxes), which could be made logistically available for computer measurement.

When this general strategy was decided upon, there were various problems which needed to be solved: The subjects have largely consisted of Wisconsin High School students who, in 1962, wrote a series of essays under controlled conditions. (There have been other subjects not so intensively studied.) There was abundant information about the Wisconsin students. The data to be analyzed for proxes consisted of various sets of essays written by these students, as key-punched literatim for computer input. The criterion for sucesss in computer strategy has consisted of the trins of expert human judges, first ratings for overall quality generated by four judges for each essay, and later ratings for ideas, organization, style, mechanics, and creativity generated by eight (different and independent) judges for each essay.

The proxes themselves consisted of various computer measurements hypothesized to have a potential relationship to the trins sought after. Some of these were statistical counts relating to length within the essay, and others were measures of types of words used. Still others investigated characteristic of sentence openings or other structures. Thirty proxes, which were most extensively explored, largely treated single words as units. Later proxes have treated various patterns of phrases, both intact and separated. All of these proxes were studied for possible correlation with the trins of essay quality, either in bivariate or multivariate relationships, and their ability to predict trins is in some ways the backbone of the empirical work to date, just as the development of the rationale, and of the various programming and statistical strategies used, is the backbone of the methodological work to date.

Findings. Chapter III specified hypotheses about certain of the proxes, and described the computer program, (called PEG, listed in Appendix A), with some of its features. Chapter IV explored the questions of reliability and validity of the proxes, and showed the ability of the computer strategy to predict the overall rating of quality about as well as the average human judge. It also discussed some of the ways in which the computer may be superior to the judge: especially in adjusting the "severity" and the dispersion of the grading system according to any uniform, predetermined standard. On two sets of essays, the computer program was able to reach multiple-regression coefficients of .71. Also, one essay's proxes were able to predict the judgments of other essays written by the same student, to a MULTR of .62. A conservative cross-validation of the program showed the ability to generate large numbers of ratings which were indistinguishable from those of the human judge. In sum, the proxes contributed significantly, in the predicted directions, to produce quite humanoid ratings of overall quality. And the Paulus tables were convenient tools for such multivariate analysis.

Chapter V made a major expansion in the program, by moving the simulation strategies to a profile of scores. The human ratings were those of 32 expert English teachers, with eight judges evaluating each of 256 essays on five major traits of writing quality, each spelled out carefully according to accepted dimensions. The individual judges were found to correlate only weakly with each other, but there was a strong tendency to a halo effect, i.e., to great uniformity of profile for any given essay judged by any given rater. However, there was a sufficient profile consensus for a singificant interaction of trait by essay. The proxes contributed differentially to the five traits and, halo aside, there were interesting relationships shown: For example, length of essay contributed highly to content, organization, and creativity, but not at all to mechanics. There was thus intuitive mutual support for the validity of the ratings and of the computer system.

The intercorrelations of the traits showed coefficients which were actually higher than the reliability of the individual traits, a surprising finding but an understandable one in view of the halo tendency, and the relative independence of the reliability. Some effort was made to cluster common judge viewpoints into a purer criterion, for purpose of simulation, and implications of this work were discussed. A most interesting comparison of this chapter was the relative ability of the computer program to simulate the various traits. Although human judges were much more reliable in judging mechanics than in judging any other trait, and somewhat less reliable in judging creativity, the computer program displayed no such handicap, and did as well with the more subjective, "qualitative" dimensions as with any.

Chapter VI made some studies of the problem of nonlinearity of prediction in such multivariate simulation. Clearly, some of the prox distributions were odd ones, and their relations with each other, and with the criterion, were irregular. The two methods of correction explored were interaction terms and transformations of the proxes. For various reasons, these were not successful in increasing the overall cross-validated multiple regression, and for practical purposes the linear assumption remained a powerful and useful one, even where it was not exactly true. Some useful programs were developed for displaying bivariate relationships and for modifying variables systematically.

Chapter VII expanded the work of the computer programming to analysis of text strings of more than one word in length. A phrase lookup algorithm was listed as an adjunct to the main program, and was used in a number of sub-studies. One of these explored the essays for the presence of standard cliche phrases. It did not find them in common or injurious use, and where they did occur their presence seemed uncorrelated with essay quality. Another substudy used the same algorithm to locate phrases believed to characterize student traits of opinionation, vagueness, or specificity. As predicted, the first two were found negatively correlated, the last positively correlated, with essay quality, but the significance could probably be accounted for by third variables of word commonness which distinguished the lists. Other substudies found null relationships between essay quality and correlative conjunctions (for one investigator) and verb voice and tense (for another). One significant study also used the phrase algorithm to examine parenthetical expressions, and found them indeed, as might be predicted, related to essay quality according to whether they occurred at the beginning (good), middle (less good), and end (perhaps poor) of a sentence. Such phrase lookup thus represented a step upwards in the power of the analysis program.

Finally, Chapter VIII implemented an on-line, interactive program to demonstrate the potential practical uses of such a system for eventual classroom applications. The program works at a time-sharing console, and is written in FORTRAN IV, like the other programs here reported. It greets the student and defines the essay assignment. When the student has finished his essay and signaled his completion, the computer (IBM 7094) begins in about 10 seconds with diagnosis, evaluation, drill, and advice. The algorithms were largely ad hoc and specific to certain narrow classes of errors. Much basic work is needed for a truly flexible system. Yet the program should help demonstrate that there is nothing in principle about the computer which will prevent a vast range of essay-analyzing applications in the future.

In short, the chapters up to this point have described the actuarial rationale, the deliberate limiting of focus, the implementation of computer algorithms, the construction of suitable criteria, and the empirical results of the current state of the art of automatic essay analysis. These chapters have also explored some statistical possibilities, various additional proxes, and some on-line token implementations in simulated settings. The remainder of this final chapter will consider certain additional possibilities of interest in the work of contemporary scholars, and will point some possible directions for the most promising future investigation of the lines here begun.

2. Some Work Related to the Project

Since the inception of Project Essay Grade, much work has gone on in areas related to the project. The investigators have made additional explorations into related disciplines, and have kept constant contact with them. For future investigators in automatic essay analysis, some knowledge of this outside but related work is essential, if they are to avoid the terrible expenses of redundance or ignorance. Therefore, this section will briefly describe some of this related work.

Journals. The related disciplines continue to grow rapidly in activity. Two journals have appeared which capitalize on the potential relevance of computation for language processing in traditional scholarship. One of these is Computers and the Humanities, since 1966 a quarterly edited by Joseph Raben at Queens College. A larger quarterly is coming out in early 1968, Computer Studies in the Humanities and Verbal Behavior, published by Mouton Press with an interdisciplinary editorial committee. (The first author here is the editorial advisor for education.) And The Journal for Educational Data Processing shows interest in natural language.

Societies. Organizationally, a great deal is happening. The Association for Educational Data Systems (AEDS) is only peripherally interested in natural language, but its involvement seems to be increasing. The Association for Computing Machinery (ACM), a very vigorous and strong organization of computer scientists numbering over 20,000, has a great deal of interest in relevant fields. It has a special interest committee for artificial intelligence (SICART), which is changing to established group status, and a group for information retrieval (SIGIR). And it has a newly forming committee for language analysis and studies in the humanities (SICLASH) which has already a substantial initial membership. The American Documentation Institute (ADI) has recently changed its name to the American Society for Information Science (ASIS), and has a keen interest in many areas overlapping this project. All of these societies put out newsletters, journals, or both. Perhaps the most acutely relevant body is the Association for Machine Translation and Computational Linguistics (AMTCL), which publishes its own journal and a useful newsletter called The Finite String. This group holds its own meetings in conjunction with ACM and the Linguistic Society of America, and has participated in two international conferences in the field.

The oldest societies within the humanities, such as the Modern Language Association (MLA), are notoriously tradition-bound, but even in the MLA a computer group is establishing a fairly permanent event at the Annual Meeting.

Besides AEDS, the educational and behavioral societies have indicated a growing interest. The pre-session training conferences held before the Annual Meeting of the American Educational Research Association (AERA) have been stimulating more sophisticated computer strategies for some years (with sponsorship from the United States Office of Education). These have been increasingly oriented toward interactive work, especially in CAI, which has strongly overlapping interests with natural-language analysis. And in 1968 we conducted the first such workshop entirely concerned with natural-language analysis for educational research.

Textbooks. A discipline has difficulty in growing rapidly until authors have defined it in suituable textbooks. There are a number of such books which bear on this work, though none is currently satisfactory for most courses which are being conceived. Works edited by Garvin (1963) and by Feigenbaum and Feldman (1964) have been mentioned earlier in this report, and so has the older one authored by Oettinger (1960) on machine translation. An excellent related work is that by Becker and Hayes (1963) on information storage and retrieval.

New arrivals include a rather descriptive book in the humanities, edited by Bowles (1967), and an important work in Automatic Language Processing edited by Borko (1967). One of the most useful works, though not readable by any but professionals, is a new book in computational linguistics by Hays (1967). A forthcoming work on computers in education, written by Allan B. Ellis, will surely feature some natural-language work. And another forthcoming work by Gerard Salton on information retrieval (due in 1968) should be valuable to some workers in natural-language analysis. A text by Veldman (1967) on FORTRAN programming for behavioral scientists, has one chapter on verbal data which should prove very useful.

In general, materials suitable for instructing in essay analysis can be pieced together from such works as these, various programming texts, works on statistics and on linguistics. But the field still lacks a suitable synthesis textbook for all introductory purposes, and work may proceed without it for some time.

Other books. On the other hand, books which have some more distant bearing on natural-language seem to be growing rapidly in number and quality, and should receive at least brief mention. In theories of automata, the growth has been especially brisk. Robert Korfhage (1966) has produced a book which relates computation to recent and current activities in mathematical logic, and the production languages described have high relevance to context free grammars and, indeed, to the basic optimism about what computers may accomplish. Marvin Minsky's book (1967) will surely open the field of computation theory to many persons who would otherwise not have made contact with it, and should thereby produce indirectly much important practical and theoretical work. And Taylor Booth (1967) has unquestionably produced the most impressive compendium on automata theory so far.

Such activity has been going on before now, but has only recently surfaced in such organized forms. In the field called "artificial intelligence," we have already seen that activity is growing with computer science. Carne (1965) has one attempted synthesis of some central concepts, and other, larger works are reportedly in preparation.

At first glance, such works may seem irrelevant to natural language processing, but the present writers do not believe that they are. Rather, they serve to change the way that computers are regarded, altering their image from that of a slavish, pedestrian worker to that of a universal machine. This seems to us a very important and necessary change in the behavioral applications of computer science.

Recent related work. Earlier portions of this report discussed some related work in other disciplines. This section will comment on some recent lines of such development, which seem particularly meaningful. This will not attempt a complete coverage of such work, but will only indicate a few of what may be major lines of related investigation, over a longer period.

We have said that the work of Project Essay has so far been actuarial in nature, leaning on statistical relations between prox and trin more than on deterministic strategies. Such statistical strategies should not be underrated. As Sapir has written, "All grammars leak." No matter how the future of such work develops, it is hard to foresee a time when serious simulation will dispense with a large probabilistic element. Yet Project Essay wishes to push ahead with the deeper linguistic and psychological dimensions as well, and to take maximum advantage of any developments in these areas.

Parsing. In the linguistic world, there are certain lines of investigation which seem very promising. One of these is in context-free and other parsing schemes aimed at syntactic analysis. Of all parsers constructed, the most realistic one so far is the Oettinger-Kuno multiple-path predictive parser at the Harvard Aiken Laboratory.

The nature of current parsing systems is described in a number of places (Garvin, 1963, pp. 223-232; Bobrow, 1967; Hays, 1967, ch. 6), and will not be explained here in any detail. Since Kuno's program enjoyed such eminence, we were very interested in possible applications, and Professor Kuno kindly arranged for parsing 50 sentences from the Wisconsin essays. The results of this processing will be briefly set forth, and illustrated.

Multiple Path System. In order to use the Kuno parsing system, every word of the text must be found in a "dictionary" -- that is, a list of words accompanied by their possible syntactic roles, encoded in a way that is useful to the system. The ordinary "noun" or "verb" is not sufficient; there are various restraints on words which are not adequately described by such broad designations, and therefore such dictionaries need painstaking construction. The Harvard dictionary is still quite limited, and some of the common student words needed to be supplied (as did all misspellings).

Figure IX-1 shows the result of looking up the words of one student sentence in this special dictionary. This sentence was: Money becomes a hindrance when it ceases to aid in the attainment of one of the best things and becomes a goal itself. Figure IX-1 shows many ambiguities in the possible syntactic roles to be played by most of the words of this sentence. Only of and and presented no homographs, and aid possessed seven homographs to compete for "the" correct parsing.

FIGURE IX-1

COMPUTER LISTING OF HOMOGRAPHS FROM
THE PARSING DICTIONARY FOR A STUDENT SENTENCE

SENTENCE NUMBER 000024 CORPUS NUMBER 01

WORD	HOMOGRAPHS						
MONEY	NNNS	MMMS	NOUS				
BECOMES	VI2S	VI3S	VT1S				
A	AAA	ART					
HINDRANCE	NNNS	MMMS	NOUS				
WHEN	IAV	CO2	RL6				
IT	TITS	PRNS	PRC				
CEASES	VT1S	VI1S					
TO	TOIS	PRE					
AID	VT1P	IT1	VI1P	II1	NNNS	MMMS	NOUS
IN	PRE	AV2					
THE	AAA	ART					
ATTAINMENT	NNNS	MMMS	NOUS				
OF	PRE						
ONE	NNNS	MMMS	NUMS				
OF	PRE						
THE	AAA	ART					
BEST	NNNC	MMMC	NOVC	AV1	AAA	ADJ	
THINGS	NNNP	MMMP	NOUP				
AND	XCO						
BECOMES	VI2S	VI3S	VT1S				
A	AAA	ART					
GOAL	NNNS	MMMS	NOUS				
ITSELF	PRO	AV1					
.	PRD						

The next illustration, Figure IX-2, shows the first parse performed by the Kuno predictive algorithm. To the scholar unfamiliar with such work, this parsing may seem a surprising example of artificial intelligence, for there is a great deal about it which would correspond with the analysis of a trained student of rhetoric. The first column is of course the list of "terminal" symbols, i.e., the words of the manifest English sentence. The second column is the "sentence structure." A little study will give some clue to the way this may be read. All words fall within the sentence "1", and we find the number 1 throughout. The word money, standing as a simple subject, is only one syntactic step from the terminal representation, and therefore we find only "1S" for structural designation. On the other hand, the word a modifies hindrance, and hindrance has the structural representation "1C" (where C stands for "complement"). Thus the article (or "adjective") a carries the designation "1CA". By such dependency relationships we have the 12-symbol depth of the and best. These words both modify things, which is the object of the preposition of, which leads the prepositional phrase which modifies the noun one, and so on back to the adverb clause headed by when, which modifies the verb becomes, the second word of the sentence. From the second column, one could thus draw a tree diagram of the sentence syntax.

The third column shows the particular syntactic category of that word for this particular parsing. A glance back to Figure IX-1 will show that all entries in this column appeared as possible homographs in the earlier output. And the fourth column is a verbal description of what that category is. The fourth column, then, depends completely on the third.

FIGURE IX-2

FIRST COMPUTER ANALYSIS OF SYNTAX
OF A STUDENT SENTENCE

NOTE: Analysis produced by the Kuno Multiple-Path Syntactic Analyzer

***** ANALYSIS NUMBER 1 SENTENCE NUMBER 000024 CORPUS NUMBER 01

ENGLISH	SENTENCE STRUCTURE	SWC	SWC MNEMONIC	SYNTACTIC ROLE	RL NUM	PREDICTION POOL
						SE
MONEY	1S	NOUS	NOUN 1	SUBJECT OF PREDICATE VERB	SENNNO	PD VSA
BECOMES	1V	VI2S	ADJ-COMPLEMENT VI	PREDICATE VERB	VXVI21	PD VSAZMNN3A
A	1CA	ART	PRO-ADJECTIVE	COMPLEMENT OF PREDICATE V	N3AAAO	PD VSAZMNN6A
HINDRANCE	1C	NOUS	NOUN 1	COMPLEMENT OF PREDICATE V	N6MMMO	PD VSAZMN
WHEN	18R	CO2	ADVERB CONJ 1	CONJUNCTION	CMCO22	PD VSACMNVZG1ZA
IT	18S	PRNS	PERSONAL PRN NOM	SUBJECT OF PREDICATE VERB	1XPRNO	PD VSACMNVSG
CEASES	18V	VT 1S	NOUN-OBJECT VT	PREDICATE VERB	VXVT11	PD VSACMNN2A
TO	18OVR	TOIS	TO FOR INFINITIVE	OBJECT INFINITIVE	N2TOIO	PD VSACMNBVF
AID	18OV	IT1	INFINITE VT1	OBJECT INFINITIVE	BVIT10	PD VSACMNN2F
IN	18OVPR	PRE	PREPOSITION	PREPOSITION	N2PREO	PD VSACMNN2FNQG
THE	18OVPOA	ART	PRO-ADJECTIVE	OBJECT OF PREPOSITION	NQAAAO	PD VSACMNN2FN5G
ATTAINMENT	18OVPO	NOUS	NOUN 1	OBJECT OF PREPOSITION	N5MMMO	PD VSACMNN2F
OF	18OVPOPR	PRE	PREPOSITION	PREPOSITION	N2PREO	PD VSACMNN2FNQG
ONE	18OVPOPO	NUMS	NUMERAL	OBJECT OF PREPOSITION	NQNNNO	PD VSACMNN2F
OF	18OVPOPOPR	PRE	PREPOSITION	PREPOSITION	N2PREO	PD VSACMNN2FNQG
THE	18OVPOPOPOA	ART	PRO-ADJECTIVE	OBJECT OF PREPOSITION	NQAAAO	PD VSACMNN2FN5G
BEST	18OVPOPOPO	NOVC	NOUN 3	OBJECT OF PREPOSITION	N5MMMO	PD VSACMNN2F
THINGS	18OO	NOUP	NOUN 1	OBJECT OF OBJECT INFINITIVE	N2NNNO	PD VSACMN
AND	1+	XCO	COORDINATE CONJ1	COMPOUND PREDICATE VERB	CMXCOO	PD VSA
BECOMES	1V	VI2S	ADJ-COMPLEMENT VI	PREDICATE VERB	VXVI21	PD N3A
A	1CA	ART	PRO-ADJECTIVE	COMPLEMENT OF PREDICATE V	N3AAAO	PD N6A
GOAL	1C	NOUS	NOUN 1	COMPLEMENT OF PREDICATE V	N6MMMO	PD
ITSELF	1D	AV1	ADVERB 1	ADVERB	PDAV10	PD
.	1.	PRD	PERIOD	END OF SENTENCE	PDPRDC	

The fifth column, however, depends also on the actual sentence structure as diagnosed by the computer program. That is, it depends on what rules of the context-free grammar, what permissible grammatical constructions, were employed in order to yield this successful parsing of the sentence. And the final columns have to do with the way the parsing is carried out, with the push-down store operating at each step of the way.

A continuing analysis of this parsing will, unfortunately, show that it does not completely match one's intuitive analysis of the later portions of the sentence. The second column indicates that becomes (fourth word from the end) is taken to be parallel with becomes (second word of the sentence). That is, it is taken to be part of a compound predicate of the word money. But most of us would take this word to be part of a compound predicate of the word it (sixth word in the sentence). The distinction, from the standpoint of "meaning," is not a trivial one at all. The way this parsing "reads" the sentence is (in reduced form): Money becomes a hindrance. . . and becomes a goal itself. Whether the distinction would be important or trivial for a particular analysis would, however, depend on the empirical situation.

A variant parsing appears in the next illustration, Figure IX-3. This was the twenty-fourth "successful" parsing of this sentence, and shows a number of changes from the first one. We see that becomes (fourth from the end) is here diagnosed as parallel with ceases, as it should be, and therefore is part of the compound predicate of it. (There is a rather subtle change in another way here, however, in the diagnosis of role of the infinitive to aid.)

FIGURE IX-3

LATER COMPUTER ANALYSIS OF SYNTAX OF
A STUDENT SENTENCE

*****ANALYSIS NUMBER 24 SENTENCE NUMBER 000024 CORPUS NUMBER 01

ENGLISH	SENTENCE STRUCTURE	SWC	SWC MNEMONIC	SYNTACTIC ROLE	RL NUM	PREDICTION POOL
						SE
MONEY	1S	NOUS	NOUN 1	SUBJECT OF PREDICATE VERB	SENNNO	PD VSA
BECOMES	1V	VI2S	ADJ-COMPLEMENT VI	PREDICATE VERB	VXVI21	PD N3A
A	1CA	ART	PRO-ADJECTIVE	COMPLEMENT OF PREDICATE V	N3AAAO	PD N6A
HINDRANCE	1C	NOUS	NOUN 1	COMPLEMENT OF PREDICATE V	N6MMMO	PD
WHEN	18R	CO2	ADVERB CONJ 1	CONJUNCTION	PDCO25	PD SGG
IT	18S	PRNS	PERSONAL PRN NOM	SUBJECT OF PREDICATE VERB	SGPRNO	PD VSG
CEASES	18V	VI1S	COMPLETE VI	PREDICATE VERB	VXVI10	PD VSGZMN
TO	18VDVR	TOIS	TO FOR INFINITIVE	ADVERBIAL INFINITIVE	CMTOIO	PD VSGCMNBVM
AID	18VDV	II1	INFINITE VI1	ADVERBIAL INFINITIVE	BVII10	PD VSGCMN
IN	18VDVPR	PRE	PREPOSITION	PREPOSITION	CMPREO	PD VSGCMNNQG
THE	18VDVPOA	ART	PRO-ADJECTIVE	OBJECT OF PREPOSITION	NQAAAO	PD VSGCMNN5G
ATTAINMENT	18VDVPO	NOUS	NOUN 1	OBJECT OF PREPOSITION	N5MMMO	PD VSGCMN
OF	18VDVPOPR	PRE	PREPOSITION	PREPOSITION	CMPREO	PD VSGCMNNQG
ONE	18VDVPOPO	NUMS	NUMERAL	OBJECT OF PREPOSITION	NQMNNO	PD VSGCMN
OF	18VDVPOPOPR	PRE	PREPOSITION	PREPOSITION	CMPREO	PD VSGCMNNQG
THE	18VDVPOPOPOA	ART	PRO-ADJECTIVE	OBJECT OF PREPOSITION	NQAAAO	PD VSGCMNN5G
BEST	18VDVPOPOPOA	ADJ	ADJECTIVE 1	OBJECT OF PREPOSITION	N5ADJO	PD VSGCMNN5G
THINGS	18VDVPOPOPO	NOUP	NOUN 1	OBJECT OF PREPOSITION	N5MMMO	PD VSGCMN
AND	18+	XCO	COORDINATE CONJ1	COMPOUND PREDICATE VERB	CMXCOO	PD VSG
BECOMES	18V	VI2S	ADJ-COMPLEMENT VI	PREDICATE VERB	VXVI21	PD N3A
A	18CA	ART	PRO-ADJECTIVE	COMPLEMENT OF PREDICATE V	N3AAAO	PD N6A
GOAL	18C	NOUS	NOUN 1	COMPLEMENT OF PREDICATE V	N6MMMO	PD
ITSELF	18D	AV1	ADVERB 1	ADVERB	PDAV10	PD
.	1.	PRD	PERIOD	END OF SENTENCE	PDPRDO	

Parsing went on and on, until there were 108 parsings of this sentence alone (a very high number in the present trials). The system has no way of automatically picking the "right" parsing from among the competitors. The knowledge about the world and about language habit which informs our own analysis has no present analogue in the serious and large-scale parsing programs. This is just the trouble with the present parsing systems, and with present linguistic knowledge, as was pointed out in an invited address by Anthony Oettinger to the 1967 Meeting of the American Documentation Institute.

But incomplete as our knowledge is, such analysis may still have much diagnostic interest and value. A great many branches of the parsing tree are pursued in such attempts, and information from these searches may have statistical value. Figures IX-4 and IX-5 show some of the statistical information which is produced by the Kuno algorithm. Such information may be useful for diagnosis of student errors, but an explanation of this possibility would take more space here than would be appropriate.

There may also be actuarial value in the ability of the program to parse any given sentence. The 50 student sentences were analyzed independently by an English scholar (Michael J. Zieky) as well as by the Kuno program, and the resulting two-way contingency layout is shown in Table IX-1. In this table the columns represent the human judgements of the 50 sentences, whether they were believed "grammatical" or "not grammatical." We see that 29 were grammatical, and 21 not so. On the other hand, the rows represent the ability of the program to find a successful parse for each of the 50 sentences. We find here that there were 29 successfully parsed, and 21 for which no parse was found. We find a very clear relation between the rows and the columns of this table. In fact, if these sentences might be assumed to be independent of one another, the resulting

FIGURE IX-4

STATISTICAL INFORMATION PRODUCED
BY THE MULTIPLE PATH SYNTACTIC ANALYZER
(SYNTAX DIAGNOSIS)

SYNTAX DIAGNOSIS

WORD NO.	ENGLISH	PATHS STILL ACTIVE	GRAMMAR SEARCHES	BLOCK TABLE SUFFICIENT	ADDRESSES INSUFFICIENT	TOTAL TEST FAILURES	EXTRA SUBRULES CONSIDERED
1	MONEY	0	3	0	2	0	11
2	BECOMES	13	42	2	1	0	2
3	A	5	10	3	0	0	0
4	HINDRANCE	3	9	4	0	0	10
5	WHEN	14	54	10	2	30	46
6	IT	28	90	18	4	4	114
7	CEASES	132	276	26	26	0	52
8	TO	104	272	204	30	134	288
9	AID	388	2716	192	192	208	672
10	IN	848	2380	1646	132	1374	2026
11	THE	2430	5780	1908	0	1086	0
12	ATTAINMENT	822	2466	724	0	62	1810
13	OF	2472	3196	2066	406	3196	4626
14	ONE	3902	11706	1098	1098	3720	4392
15	OF	2868	3860	2258	610	3174	5284
16	THE	4978	9956	2502	0	1802	0
17	BEST	700	4200	1960	0	938	2520
18	THINGS	3542	15858	1848	600	5540	7644
19	AND	4552	6128	2254	28	836	0
20	BECOMES	1446	4338	296	112	242	296
21	A	462	924	288	0	150	0
22	GOAL	138	414	276	0	612	690
23	ITSELF	354	828	306	0	198	0
24	.	108	108	108	0	0	0
	TOTAL	30309	75614	19997	3243	23306	30483

FIGURE IX-5

FURTHER STATISTICAL INFORMATION PRODUCED
BY THE MULTIPLE PATH SYNTACTIC ANALYZER
(SYNTAX SUMMARY)

SYNTAX SUMMARY FOR SENTENCE NUMBER 000024 CORPUS NUMBER 01

SUMMARY OF PATH ELIMINATING TEST FAILURES

TYPE OF TEST	NUMBER OF FAILURES
POOL OVERFLOWS	0
SHAPER OVERFLOWS	9108
NESTER OVERFLOWS	4198
NUMBER AGREEMENT	0
CN	4518
XC/XD	868
CN/CM/XC/XD	4302
PA	312
SELF-EMBEDDING	0
COMPOUND COMPATIBILITY	0

START TIME 0.0
END TIME 0.0

0.0

TABLE IX-1

THE RELATION OF COMPUTER PARSING TO
JUDGED GRAMMATICALNESS OF STUDENT SENTENCES

Machine Parsing	Human Judgment "Grammatical"	"Not Grammatical"	Row Sums
Parsed	24	5	29
Not Parsed	5	16	21
Column Sums	29	21	50

NOTE: All machine parsing was done through the courtesy of S. Kuno, Harvard University.

Chi square = 15.04 ($p < .001$)*

Contingency coefficient = .48

*Data were not independent. See discussion in text.

chi square of 15.04 would be significant beyond the .001 level of confidence. And the related contingency coefficient would be a healthy .48. In other words, the ability of the program to parse a sentence would have some predictive power for whether the sentence would be judged grammatical by an expert human. Because of the casual way these sentences were drawn for computer analysis, the assumption of independence is not warranted; but the general trend of the results still suggests actuarial value in the use of such algorithms for computer analysis of essays.

The data from the comparison are presented in a different way in Table IX-2. Here we are able to review the computer analysis of the sentences. Ideally, of course, every sentence should produce only one parse, and that one should be the same as that of an expert human. Nevertheless, it is important that those sentences which were grammatical had, on the average, many <u>more</u> completed parsings than those which were not grammatical. And it is interesting that the median number of parsings for grammatical sentences was 3, but 0 for the ungrammatical ones.

It is also interesting to observe, in Table IX-2, the order in which the correct parsings occurred. Only 16 parsings were judged as intuitively faultless. Seven of these occurred on the 1st trial, 6 on the 2nd, and the others as shown. The present Kuno program, outstanding as it is, has made no provision for statistical optimization, and this performance should be improvable in some appropriate adaptation.

In order to have a similar parser for experimental purposes, we have undertaken to make a PL/I version of the predictive parser, programmed for the Project by Gerald Fisher, and listed in Appendix D. Appendix D also has the flowchart of that parser, which may help the reader new to such strategies to understand their nature. This program,

TABLE IX-2

MACHINE PARSING PERFORMANCE OF GRAMMATICAL AND UNGRAMMATICAL STUDENT SENTENCES

	Human Judgment	
	"Grammatical"	"Not Grammatical"
N Sentences	29	21
Mean of Machine Parsings	23.68	8.14
Median of Machine Parsings	3	0
N Parsings Judged Correct	16	
Order of "Correct" Machine Parsings	7 on 1st 6 on 2nd 1 on 3rd 1 on 38th 1 on 52nd	

NOTE: All machine parsing was done through the courtesy of S. Kuno, Harvard University.

called PARSE, has been debugged and tested with artificial information, but not yet with natural language. Such a parsing program is only the vehicle; the content must be furnished by (1) a suitable dictionary; and (2) a suitable set of grammatical rules. This brief chapter is not the place to set down all the considerations which will play a role in any further development of such linguistic processors, but a few points will be suggested by the next sections.

Discourse analysis. Further pondering of the sentence parsing will reveal some difficulties not considered by the multiple path analyzer. If one is concerned about "meaning" and about how a machine "reads" a sentence, then one must arrange for the prose of an essay to hang together, in some sort of cognitive net. The token sentence of Figure IX-2 will illustrate this problem. No provision is made for the analysis of antecedents or referents: the pronoun it is not tied in any mechanical way to the word money which it presumably renames. But pronouns are not the only offenders in such a simplified analysis. In most prose, such as scientific writing, a large proportion of the nouns refer in some abbreviated way to persons, objects, or ideas which have already been treated in the writing. The human reader at once connects these new expressions with those which have gone before, but how this is accomplished is not yet understood very clearly.

J. Olney and D. Londe, of the System Development Corporation, are among the very few who have given computational attention to this problem, and their brief writings are not yet ready for any broad dissemination (personal communications). There are clearly some explicit cues which may be helpful (such as number, gender, person). There are synonym relationships also, some of which may be discovered through mechanical use of a large dictionary.

There are also questions of proximity; other things being equal, one would expect the most recent candidate for referent to be operative. Standard techniques of optimization may weight such criteria appropriately and may make a best-guess selection of reference for pronouns or other anaphoric expressions. A great deal of work is necessary, then, in this field of discourse analysis.

Transformational grammar. Of course, one of the most active areas for current linguistic research is in transformational grammar. Treatments of this topic may be found in a number of references (e.g., Hays, 1967, ch. 8). Perhaps the best recent treatment of the topic, especially from the viewpoint of computation, is by Keyser and Petrick (in press), both of whom have served as consultants for Project Essay. Perhaps the most useful program for transformational analysis is that described in Petrick's thesis (1965).

John Moyne and David Loveman, at the IBM Boston Programming Center, have programmed a very limited system which carries analysis through a syntactic analysis to a transformational analysis, and prints out appropriate answers to questions. Like all such extant systems, this one is for a special purpose, in this case document retrieval from a large library. And they have processed a few student sentences, from Project Essay, on an experimental basis, through their first, surface-structure parser.

Semantics. In general, transformational grammars are far from any linguistic perfection, and face deep problems which will not be described here. Yet there are approaches to the question of meaning which have some demonstrable usefulness and power, and which may sidestep these deepest problems for the purposes of application. Some of these are generally described as involved with "semantics," and some are framed within the practical problem of question-answering systems. Still others are spoken of in terms of information storage and retrieval, especially what is spoken of as "fact retrieval," as contrasted with "document retrieval".

These works share a common concern with the way that information may be read into some data representation in the computer, and how it may then be made accessible for further use. William A. Woods (1967), for example, took for granted the output from some syntactic and transformational parsing system, and then asked how he could develop a question-answering system. His particular corpus was flight information from the Airlines Guide, and he worked out operators for logical comparison and other semantic concerns which would implement such a system. In doing so, he built upon earlier work with BASEBALL (see Feigenbaum and Feldman, 1963, Sec. 5), and similar systems, but went beyond his predecessors in certain important ways. Other new work is that of Quillian (1966), who has provided a way of storing semantic relationships. His structures permit comparison between two statements, and make possible judgments about them concerning their agreement, disagreement, or irrelevance.

The importance of symbolic logic in such systems is apparent in the recent work by Levien and Maron (1967). These authors use the predicate calculus, with binary relations only, as a universal tool for fact storage. They organize a data base which has four different ways of random access (corresponding to sentence number, relation name, and the two elements) for rapid retrieval of the fact through any of its components. Their method is wasteful of storage space, but extremely rapid in operation, able to locate any fact without poring through lists. Their system thus enjoys some important virtues of the psychological models.

There is much work going on, then, in fields with important relevance for the future of essay analysis. It takes the form of progress in linguistics, psychology, and computer science, and elements of statistics and logic have a bearing as well. Surely, Project Essay must maintain its close contacts with these fields in relation to its future work.

3. Future Work in Essay Analysis

Need for Flexibility. The sub-discipline of computer analysis is only now beginning to take shape. In the meantime, as we have seen, work in the area seems to call for a rather unusual approach: interdisciplinary, broad in purpose, and flexible.

In its present development, the computer analysis of essays does not yet lend itself to the clear, Fisherian, "classical" experimental designs, because not all operations can be foreseen. It does, however, permit clear procedures of dynamic development and exploration at each stage of the study, and verification of accomplishment at the end. Properly understood, these characteristics are not handicaps, but symptoms of large research scale. In a recent paper, Baker (1965) pointed out that the larger and more exploratory research project "must be inherently dynamic and possess the ability to change its internal structure without sacrificing the rigor of the design" (p. 15). And another writer (Doyle, 1965) has recently stated that as a study approaches the "basic research end of the spectrum, it becomes more and more imperative to be free to alter the plan. Indeed, in basic research altering the plan ought to be a state of mind." With the present study, it would be mistaken and even misleading to commit the investigation prematurely to too narrow a path.

The first phases of this study illustrate this point. In the earlier work, only the most general goal then, as now, was completely operational, foreseeable, and attainable: the maximization of the correlation between computer-analyzed prose characteristics and the human judgments of the prose. The earlier work has reached this goal (so far as possible during the time permitted), but many paths were altered along the way. Programming plans were modified and improved. Certain hypotheses were reformulated, and others discarded. At the conclusion of the first phases, the progress has been much greater than if the inevitable misconceptions of the beginning had been adhered to in spite of everything. The ultimate goal, however, was rigorously adhered to, and the most careful investigatory techniques employed at each decision point along the way. In the newer research designs, what must be done, rather than to make all of the decisions before the choice points are reached, is to illustrate the quality of decision-making. This portion of the proposal, and that which follows, are intended to state the general objectives and the decision-making strategy by which these goals will be attained.

As noted before, the work reported here has already identified useful computer-analyzable indicators of student writing skill, and has demonstrated the potential feasibility of overall theme evaluation by computers. When holistic grades are desired, or ratings of important essay traits, the PEG computer program already assigns marks as accurately (measured against the criterion of multiple expert judgments) as the individual, trained English teacher. Future work should expand the work to the analysis and evaluation of content, and deepen it linguistically and psychologically by investigation of more humanoid processes. Some general future objectives may be outlined:

1. To expand consideration to essay content as well as style.

2. To explore the relation of dictionary strategies to successful analysis, and to develop optimum strategies for the Random House Dictionary tape.

3. To analyze computer-generated data in relation to subjective measures of content and style in the early secondary years, to increase usefulness of analysis.

4. To improve the programming of on-line correction of essays, and on-line feedback to the student or teacher.

5. To identify future strategies for deeper exploration of this new field of educational technology.

Grading of content. Just as we have opened up the possibility of grading the esthetic traits of an essay in English, so we should also be interested in the possibility of judging the substantive content of essay material, apart from the general writing ability of the student. This is a dimension of essay analysis not yet attempted within this project, yet it may be approached at a number of different levels of sophistication, and some of these might prove both economical and rewarding. Let us consider a sample problem in American history, to conceptualize these various levels, first heuristically, and finally in more hypothetical but technical detail.

Suppose we wished to grade children on the factual content of an essay about the discovery of America. It might be supposed that certain words or phrases should appear in the more complete essays: Columbus, Christopher, Ferdinand, Isabella, king, queen, Spain, Azores, 1492, Nina, Pinta, Santa Maria, Indians, etc. These words and others could be fed into core as a kind of dictionary, much as has been done already with such lists as prepostions, misspellings, common words, etc. Each first use of any of these Columbus expressions could be scored in some fashion. No doubt such scores would be positively correlated with "factual completeness" ratings as assigned by human judges. Such scoring would therefore be an aid in achieving the simulation sought for in Quadrant I.A of Figure II-1.

Suppose we asked for meaningful relationships among these and other words. One evidence of such a relationship might be to have the word Isabella occur in the same sentence as the phrase queen of Spain. And such use within the same sentence should perhaps receive a higher score than use in different sentences. Again, the consideration is actuarial, yet now the statistical analysis is one small step closer to a meaningful relationship between ideas.

Of course, at a somewhat higher level, we would not wish too high a premium placed on arbitrary words, so we might include monarch or sovereign in core storage as acceptable equivalents of queen, or Isabel as an acceptable form of Isabella. Synonyms could possibly be scored quite sensitively, according to their judged "semantic distance" from the most desired words.

Or we could look further for meaning, by asking that, within the sentence, Isabella and queen (or equivalents) be in some standard form suggesting identity. Some common ways this might be done are as a title (Queen Isabella), as an appositive (Isabella, queen . . . or queen . . . . Isabella), or as a predicate nominative (Isabella . . . [form of to be] . . . queen, or inverse). And such evidence of identify could be scored somewhat more highly than the appearance together without such evidence.

Now consider a much more advanced system. Note that if we have a sufficiently sophisticated general dictionary available, and an adequate general sentence analyzer, we will not need to anticipate each specific equivalent expression or relationship in each specific essay examination, in order to score it. We can instead read in a key in the form of English sentences containing some model narrative about Isabella and Columbus. Various equivalences would then be potentially available for the grading of the student's "own words." But here we are clearly in the I.B Quadrant of Figure II-1, and are doing a kind of "master analysis."

The progress here is from the employment of a simple lexicon of key words, to the acceptance of their synonyms, to the search for the key words or synonyms in appropriate contexts, to the search for these meanings in appropriate relationships.

For any ultimate, applied evaluation of essay content, a computer program should be no more elaborate than necessary for the overall goal. If it is much cheaper to use the lower, lexical strategy, and if it is almost as accurate, then it would only waste machine time to compute higher-level information which will not be used. On the other hand, in some essays the special vocabulary may not be very important, and other factors may control the evaluation of merit. The Columbus example seems to depend highly on vocabulary, but there may be others in which all students use essentially the same special vocabulary, and the discriminations are at the higher contextual and relational levels.

One early discovery needed, then, is the degree to which most school essay evaluation is dictionary-loaded. And some workers are addressing themselves to this need, with some college level examinations, at the time of writing.

Another purpose is to seek more advanced strategies of semantic analysis, of the contextual or relational sort. These strategies have some antecedents as well. Most techniques of informative retrieval, for example, are based upon co-occurrences (cf. pp. 310-353 in Garvin, 1963). And the usual employment of the General Inquirer system employs such contextual techniques (Stone, 1966). As we have said, still more advanced systems of relational semantic analysis have been programmed by such workers as Woods (1967), or John Moyne (of IBM's Boston Programming Center). An impressive attack on the problem of artificial memory appropriate to such relationships has been made by M. Ross Quillian (1966). These workers have already consulted informally with this project, and will be available for further work, and some have already done some pilot tasks.

Further analysis of style. Surely workers should not abandon the analysis of esthetic quality in writing, but rather use available advanced strategies of meaning to further such judgment. To some extent, the analysis of style must be paired with meaning, and it is hoped that the next years will see advances in the description of surface structure, and at least some preliminary consideration of stylistic traits such as synonym, contrast, and parallelism, all of which have a large semantic content.

On a tentative basis, as we have described, some high school essays have been already analyzed by parsing programs, one written in FAP for the IBM 7094 by Susumu Kuno (1964) at Harvard, and the other in PL/1 "ELF" by David Loveman and John Moyne at the Boston Programming Center. These have indicated that partial parsing is already available, but that further adaptation of any parser will be necessary. To some extent, the output of a parser will be used to inform the semantic analysis. During the next years of such work parsing will be carried much further than at present.

Hypothetical complete essay analyzer. Some reasonable future objectives of workers have been stated above. These are realizable and useful objectives, and can probably be obtained within reasonable limits of time and effort. Nevertheless, it is informative to construct a more distant objective, which would be a set of computer routines tied together in a more complete and humanoid essay analyzer.

Anticipated future strategies are currently summarized in Figure IX-6. This figure is based partly on work already accomplished, partly on suggested minor adaptations of systems already working for others, and partly on projected programs which are not yet operative in any system, but which do not seem impossibly difficult at the efficiency desired.

FIGURE IX-6

HYPOTHETICAL COMPLETE ESSAY ANALYZER

1. INPUT and PUNCH. Handwritten or typewritten or other raw response of the writer is converted for computer input.

2. SNTORG. Creates arrays of words and sentences as found in prose. This is just as performed in PEG, or by a PL/I version called SCORTXT.

3. DICT. Assignment of available syntactic roles to each word. This is currently done by many programs, but needs an expanded dictionary, and ambiguity resolver. At the same time, the semantic information will be stored in the work-space for reference of other parts of program. The tape-written Random House Dictionary (Unabridged) is a very valuable facility for this work.

4. PARS. A modified Kuno (1964) program such as PARSE seems most promising, and the skeleton is now available in PL/I. Alterations will be necessary to accept well-formed substrings, and to work out dictionaries and grammars of appropriate power.

5. REFER. This is intended to identify and encode the most likely referents of pronouns and other anaphoric expressions. This process must employ both syntactic features and probably semantic information from DICT or other sources.

6. KERNEL and STRUC. From the rewritten string output of (5), KERNEL would extablish a set of elementary propositions, and STRUC would encode the relationships among these elements. This step would retain the information of an essay in simplest possible units, yet would retain additional information about emphasis, subordination, causal relation, etc., among these units.

7. EQUIV. The elementary units would be augmented by the semantic information in DICT. To each word would be assigned a cluster of permissible synonyms, with weightings of semantic distance. This permits an analysis of redundance and emphasis in the essay, and permits a comparison of the content of the student essay with that of the key or master essay.

8. STYLE. Descriptions of the surface structure characteristics of the essay: parts of speech, organization of themes, types and varieties of sentence structure, grammatical depths, tightness of referience, etc.: information about grammatical errors and strengths.

9. CONTNT. Comparison of the agreement of student and master essay, through measure of kernel hits and struc hits, these weighted by semantic distance of language chosen.

10. SCOR. Multivariate prediction of appropriate profile for the immediate purpose.

The limitations of space will permit only a few comments on this figure. For large grading systems, over established substantive content, it would be possible, for the key or master essay, to edit by hand the output from certain routines (especially REFER and STRUC). Of course, four of the most important routines listed in Figure IX-6 are far from perfected in any existing programs. Ideally, they would assume better solutions to certain major, stubborn problems in computational linguistics.

Indeed, certain steps in this hypothetical essay grader are close to the heart of some of the most persistent and troublesome problems in linguistics. Is it necessary that sentences be syntactically analyzed before mapping into deep structure? What is the proper role of semantics in such deep structure? How can the outside knowledge of the reader be incorporated into the machine analysis? In general, how may we incorporate some of the intuitive richness which the literate human brings to his reading?

Surely, in essay analysis workers will not suddenly resolve all such questions. These questions so trouble linguists as to contribute to the recent official pessimism, in the United States, about the future of mechanical translation. After 15 years of effort, mechanical translation is still regarded as disappointing in quality, and virtually no sustained output of any machine program would be ordinarily mistaken for the work of a professional human translator.

On the other hand, the earliest attempts at essay grading by computer have, in a very limited way, leaped ahead of machine translation. And if the expert human ratings of high school essays may be regarded as an acceptable goal, then the machine program appears to have reached such a goal already. For that matter, improved performance, even superior to that of the individual human expert, appears to be immediately practicable as well.

The explanation of this advantage, of course, is that the problem of essay grading as attacked in the current work is much easier than the problem of machine translation. In translation, every nuance of the input string should be accounted for in the output string. In essay grading, only a certain portion of the input text needs to be accounted for, and the output does not depend on the existence of any large language-generating system. High quality machine translation apparently demands a fair portion of the total language-manipulating capability of the human, but essay grading may use only a fraction of it, and may process language in ways quite different from that of the human being. For example, our present programs have to date largely ignored order and sequence in the essays, although to the human the order of words is, of course, of crucial and unceasing importance.

Since essay grading can work with such fractional information, then, why pursue the deeper analysis of Figure IX-6? Clearly, the purpose is not entirely the same as it would be for the usual linguist. At any discrete time in research, what is sought is not necessarily the perfect humanoid behavior, but rather those portions of that behavior which, given any current state of the art, will contribute optimally to efficient and practicable improvements in output. Indeed, regardless of the eventual perfection of deep linguistic behavior, for any specific application to essay analysis, at any one moment, large portions of such available behavior may be irrelevant, just as it seems that ordinary human language processing does not usually call for our full linguistic effort.

Yet we regard it as eventually important to be able to perform these various kinds of advanced machine analysis when required. Therefore, the eventual uses of the ideal essay analyzer may require analytic capability as deep as may be imagined. Writing out suitable comments for the student, for example, will in some cases tax any system which may be foreseen.

Even approximate solutions to these problems, however, though unsatisfactory for certain scientific purposes, could make important contributions to the educational description and evaluation of essays. For such evaluation is itself probabilistic, limited by imperfect asymptotes of writer consistency and rater agreement. And such evaluation therefore does not require, to be practicable and satisfactory, a deterministic perfection. There is a fundamental difference in goals which must be realized. As has been demonstrated here, the output from much cruder statistical programs has already reached a quality not too remote from usefulness. The more advanced strategies currently seem, at least to the present workers, bright with promise, for an ultimate target of such analysis, subject to alteration and amendment as more is learned about the nature of essays and about the evaluative process.

In conclusion, this section on the future has aimed, first, at explaining the special nature of objectives in a new, exploratory, and developmental research; second, at briefly listing concise and obtainable objectives; third, at explaining appropriate goals in the evaluation of subject-matter content and in the appropriate use of dictionaries; fourth, at explaining the relation between objectives of stylistic analysis and objectives of subject-matter; and fifth, at setting forth ultimate objectives in a humanoid, hypothetical analyzer which, while it will never be <u>completely</u> realized, will be a target for the accomplishment of the immediate future. Surely, the computer analysis of language will become a permanent feature of the educational scene.

APPENDIX A

```
D-445    PEG I          FORTRAN SOURCE LIST                09/26/66     PAGE 1
ISN   SOURCE STATEMENT
0 $IBFTC PROJECT ESSAY
  C
  C
  C   THIS IS THE COMPLETE SOURCE PROGRAM FOR ESSAY ANALYSIS, 1966-67,
  C   SUPPORTED BY THE COLLEGE ENTRANCE EXAMINATION BOARD,
  C   AND MORE BY THE UNITED STATES OFFICE OF EDUCATION
  C   AND CARRIED OUT AT THE BUREAU OF EDUCATIONAL RESEARCH AT THE UNIVERSITY
  C   OF CONNECTICUT, IN STORRS, WHERE ELLIS B. PAGE WAS DIRECTOR OF THE
  C   PROJECT, AND MR. AND MRS. GERALD FISHER WERE PRINCIPAL PROGRAMMERS.
  C
  C      IN THE EARLY WORK, HIGH SCHOOL ESSAYS WERE ANALYZED BY
  C   COMPUTER FOR 30 VARIABLES.  THESE VARIABLES ARE
  C   TRANSFORMED BY THE SAME PROGRAM TO APPROPRIATE SCALES (USUALLY
  C   RATIOS).  THEN MULTIPLE REGRESSION MAY BE PERFORMED TO PREDICT THE
  C   POOLED HUMAN JUDGMENTS OF THESE ESSAYS.  THE PRESENT PROGRAM
  C   DOES THE CENTRAL TASKS OF SENTENCE ORGANIZATION AND WORD LOOKUP
  C   THAT ARE IMPORTANT IN ALMOST ANY NATURAL-LANGUAGE ANALYSIS.
  C
  C      ADAPTIONS OF THIS PROGRAM MAY BE MADE RATHER EASILY.
  C   INQUIRIES MAY BE ADDRESSED TO DR. PAGE AT THE BUREAU OF EDUCATIONAL
  C   RESEARCH, UNIVERSITY OF CONNECTICUT.
  C
  C      THIS IS THE MAIN PROGRAM, CONTAINING
  C   THE GROSS LOGIC FOR PROCESSING THE ESSAYS.
  C   THE BEGINNING STATEMENTS OF EACH PROGRAM SPECIFY THE INTERRELATION-
  C   SHIPS AMONG THE PARTS OF STORAGE AND WORD TYPES AND WORD NAMES.
  C   THESE STATEMENTS ALLOW FLEXIBILITY IN WORD HANDLING, COMPRESSION
  C   IN PROGRAMMING, AND THE NECESSARY COMMUNICATION LINKS BETWEEN
  C   THE INDIVIDUAL SUBPROGRAMS.
  C   THE WORDS, THEIR CONTENTS, THEIR ALTERNATIVE NAMES AND THEIR
  C   LOCATIONS IN STORAGE ARE DEFINED ON AN ACCOMPANYING ALPHABETIZED
  C   LIST.
  C
  C
  C      BESIDES THE TABLES OF CHECKLIST WORDS, INPUT CONSISTS OF ESSAYS
  C   WHICH ARE PUNCHED ONE LINE PER CARD, USING UP TO 80 COLUMNS OF THE
  C   CARD.
  C   THE FIRST CARD OF EACH ESSAY IS PRECEDED BY AN IDENTIFICATION CARD
  C   WHICH CONTAINS THE IDENTIFICATION NUMBER OF THIS ESSAY IN COLUMNS 1-5
  C   AND THE TITLE INDICATOR (WHICH IS BLANK IF NO TITLE IS PRESENT).
  C   FOLLOWING THE LAST CARD OF EACH ESSAY IS THE END CARD WHICH
  C   CONTAINS AN ASTERISK (*) IN COLUMN 1, AND A BLANK IN COLUMN 2.
  C   FOLLOWING THE LAST END CARD IS THE END OF JOB CARD WHICH
  C   CONTAINS 99999 IN COLUMNS 1-5.

  C
  C      THE OUTPUT CONSISTS OF PRINTED LINES, ONE FOR EACH SENTENCE, AND
  C   AN ADDITIONAL ONE FOR EACH ESSAY CONTAINING IN ARRAY
  C   ORDER THE CONTENTS OF THE SUMS ARRAY AND THE TOT ARRAY.
  C   THESE CONTENTS ARE DESCRIBED ON THE ALPHABETIZED LIST.
  C   FOLLOWING IS THE SET OF TRANSFORMATIONS, FROM WHICH
  C   THE REGRESSION ANALYSIS MAY BE RUN.
  C   THE SUMMARY ESSAY DATA ARE ALSO PUNCHED IN CARDS FOR IMMEDIATE USE.
C
C   WE WRITE THE PRINTED INFORMATION ON TAPE UNIT 0
```

WE WRITE YHE TOTAL INFORMATION ON TAPE UNIT 4

READ THE WORD LISTS AGAINST WHICH PARTS OF EACH ESSAY WILL BE CHECKED

NAME	COMMON	TYPE	ARRAY NAME	CONTENTS
A	CHAR	REAL	ALPHA% 1<	ABBBBB
APOSTR	CHAR	REAL	PUNCT% 6<	BBBBB
B	CHAR	REAL	ALPHA% 2<	BBBBBB
BLANK	CHAR	REAL	PUNCT% 1<	BBBBBB
BROKUP	IN	REAL	BROKUP	80 WORD CARD IMAGE 1 COLUMN PER W
C	CHAR	REAL	ALPHA% 3<	CBBBBB
CONNEC	LISTS	DOUBLE	RDTBL%162<	30 CONNECTIVES
COMMA	CHAR	REAL	PUNCT% 4<	,BBBBB
COLON	CHAR	REAL	PUNCT%11<	..BBBB
CPAREN	CHAR	REAL	PUNCT% 8<	<BBBBB
D	CHAR	REAL	ALPHA% 4<	DBBBBB
DALE	LISTS	DOUBLE	RDTBL%222<	3000 DALE LIST WORDS
DASH	CHAR	REAL	PUNCT%15<	--BBBB
DECLAB	LISTS	DOUBLE	RDTBL%10222<	10 WORDS TO IDENT DECLA B
DECPT	CHAR	REAL	PUNCT% 3<	.BBBBB
E	CHAR	REAL	ALPHA% 5<	EBBBBB
ENDPCT	OUT	INTEGER	SUMS%28<	1 OR FOR PUNCT AT END OF SENTENC
EXCLAM	CHAR	REAL	PUNCT%13<	.XBBBB
F	CHAR	REAL	ALPHA% 6<	FBBBBB
G	CHAR	REAL	ALPHA% 7<	GBBBBB
H	CHAR	REAL	ALPHA% 8<	HBBBBB
HLFTXT	IN	REAL	REAL STORAGE OF CURRENT SENTENCE	
HYPHEN	CHAR	REAL	PUNCT% 5<	-BBBBB
I	CHAR	REAL	ALPHA% 9<	IBBBBB
ID	OUT	INTEGER	SUMS% 1<	IDENT NO THIS ESSAY
ITALIC	CHAR	REAL	PUNCT%16<	%/<BBB
J	CHAR	REAL	ALPHA%10<	JBBBBB
K	CHAR	REAL	ALPHA%11<	KBBBBB
L	CHAR	REAL	ALPHA%12<	LBBBBB
LENGTH	IN	INTEGER	LENGTH	1-12 FOR WD LENGTH 99 FOR PUNCT T
M	CHAR	REAL	ALPHA%13<	MBBBBB
N	CHAR	REAL	ALPHA%14<	NBBBBB
NAPOS	OUT	INTEGER	SUMS%11<	NO OF APOSTROPHES THIS SENTECNE
NCOMMA	OUT	INTEGER	SUMS%12<	NUMBET OF COMMAS THIS SENTENCE
NCOLON	OUT	INTEGER	SUMS%17<	NO OF COLONS THIS SENTENCE
NCONN	OUT	INTEGER	SUMS%23<	NO OF CONNECTIVES THIS SENTENCE
NDASH	OUT	INTEGER	SUMS%16<	NO OF DASHES THIS SENTENCE
NDALE	OUT	INTEGER	SUMS%27<	NO OF DALE WORDS HTIS SENTENCE
NEXCLA	OUT	INTEGER	SUMS%20<	NO OF EXCLAMATION PTS THIS SENCEN
NPAREN	OUT	INTEGER	SUMS%10<	NO OF PARENTESES THIS SENTENCE
NPER	OUT	INTEGER	SUMS%13<	NO OF PERIODS THIS SENTENCE
NPERCT	OUT	INTEGER	SUMS%14<	NO OF PERENCT SIGNS THIS SENTENCE
NPREP	OUT	INTEGER	SUMS%22<	NO OF PREPOSITIONS THIS SENTENCE
NQUOTE	OUT	INTEGER	SUMS%19<	NO OF QUOTES THIS SENTENCE
NQUES	OUT	INTEGER	SUMS%21<	NO OF QUESTION MARKS THIS SENTENC
NRELPR	OUT	INTEGER	SUMS%25<	NO OF RELATIVE PRONOUNS THIS SENT
NSEMIC	OUT	INTEGER	SUMS%18<	NO OF SEMICOLONS THIS SENTENCES
NSPELL	OUT	INTEGER	SUMS%24<	NO OF SPELLING ERRORS THIS SENTEN
NSCONJ	OUT	INTEGER	SUMS%26<	NO OF SUBORDINATING CONJUNCTIONS

NUMSEN	OUT	INTEGER	TOT% 3<	NO OF SENTENCES THIS ESSAY
NUMPAR	OUT	INTEGER	TOT% 4<	NO OF PARA THIS ESSAY
NUMWDS	OUT	INTEGER	SUMS% 8<	NO OF WORDS THIS SENTENCE
NUNDER	OUT	INTEGER	SUMS%15<	NO OF ITALICIZED WORDS THIS SENTE
NWDSQ	OUT	INTEGER	SUMS% 9<	SQ OF NO OF WDS THIS SENTECE
O	CHAR	REAL	ALPHA%15<	OBBBBB
OPAREN	CHAR	REAL	PUNCT% 7<	%BBBBB
P	CHAR	REAL	ALPHA%16<	PBBBBB
PARNUM	OUT	INTEGER	SUMS% 4<	SEQ NO OF THIS PARAGRAPH
PERIOD	CHAR	REAL	PUNCT%10<	.BBBBB
PREP	LISTS	DOUBLE	RDTBL%42<	50 PREPOSITIONS
Q	CHAR	REAL	ALPHA%17<	QBBBBB
QUEST	CHAR	REAL	PUNCT%14<	.QBBBB
R	CHAR	REAL	ALPHA%18<	RBBBBB
RDTBL	LISTS	REAL	RDTBL	10540 WORDS CONATIN WORD TABLES
RELPRO	LISTS	DOUBLE	RDTBL%142<	10 RELATIVE PRONOUNS
S	CHAR	REAL	ALPHA%19<	SBBBBB
SENNUM	OUT	INTEGER	SUMS% 3<	SEQ NO THIS SENTENCE
SENTYP	OUT	INTEGER	SUMS%29<	1 IF DECLAR A 0 IF NOT
SENTYP	OUT	INTEGER	SUMS%30<	1 IF DECLAR B, 0 IF NOT
SENTYP	OUT	INTEGER	SUMS%31<	1 IF EXCLAM, 0IF NOT
SENTYP	OUT	INTEGER	SUMS%32<	1 IF QUESTION, 0 IF NOT
SEMIC	CHAR	REAL	PUNCT%12<	.,BBBB
SLASH	CHAR	REAL	PUNCT% 9<	/BBBBB
SPELLX	LISTS	DOUBLE	RDTBL%6222<	2000 COMMONLY MISPELLED WORDS
SSQLET	OUT	INTEGER	SUMS% 7<	SUM OF SQ OF LETTERS BY WORD THIS
STAR	CHAR	REAL	PUNCT% 2<	*BBBBB
SUBVER	OUT	INTEGER	SUMS% 5<	1 FOR S-V TYPE OPEN 0 FOR NO
SUMLET	OUT	INTEGER	SUMS% 6<	SUM OF THE LETTERS THIS SENTENCE
SUBCON	LISTS	DOUBLE	RDTBL%2<	20 WDS FOR SUB CONJ TEST
SVOPEN	OUT	INTEGER	TOT% 5<	NO OF SENT OPENING S-V
SVOPN	LISTS	DOUBLE	RDTBL%10242<	150 WORDS FOR S-V OPEN TEST
T	CHAR	REAL	ALPHA%20<	TBBBBB
TAPOS	OUT	INTEGER	TOT%11<	NO OF APOSTROPHES THIS ESSAY
TCOMMA	OUT	INTEGER	TOT%12<	NO OF COMMAS THIS ESSAY
TCOLON	OUT	INTEGER	TOT%17<	NO OF COLONS THIS ESSAY
TCONN	OUT	INTEGER	TOT%23<	NO OF CONNECTIVES THIS ESSAY
TDASH	OUT	INTEGER	TOT%16<	NO OF DASHES THIS ESSAY
TDALE	OUT	INTEGER	TOT%27<	NO OF DALE WORDS THIS ESSAY
TEXCLA	OUT	INTEGER	TOT%20<	NO OF EXCLAMATION PTS THIS ESSAY
TENDPT	OUT	INTEGER	TOT%28<	NO OF SENT WITH NO END PNCT THIS
TEXT	IN	DOUBLE	HLFTXT	ASSEMBLED WORDS OF THIS SENTENCE
TID	OUT	INTEGER	TOT% 1<	IDENT NO THIS ESSAY
TITLE	OUT	INTEGER	SUMS% 2<	U IF YES TITLE,0 IF NO TITLE
TOTLET	OUT	INTEGER	TOT% 6<	SUM OF LETTERS THIS ESSAY
TOTWDS	OUT	INTEGER	TOT% 8<	NO OF WORDS THIS ESSAY
TPAREN	OUT	INTEGER	TOT%10<	NO OF PARENTHESES
TPER	OUT	INTEGER	TOT%13<	NO OF PERIODS THIS ESSAY
TPERCT	OUOUT	INTEGER	TOT%14<	NO OF PERCENT SIGNS THIS ESSAY
TPREP	OUT	INTEGER	TOT%22<	NO OF PREPOSITIONS THIS ESSAY
TQUOTE	OUT	INTEGER	TOT%19<	NO OF QUOTES THIS ESSAY
TQUES	OUT	INTEGER	TOT%21<	NO OF QUESTION MARKS THIS ESSAY
TRELPR	OUT	INTEGER	TOT%25<	NO OF RELATIVE PRONOUNS THIS ESSA
TSQLET	OUT	INTEGER	TOT% 7<	SUM OF SQ. LETTERS IN EACH WORD
TSEMIC	OUT	INTEGER	TOT%18<	NO OF SEMICOLONS THIS ESSAY
TSPELL	OUT	INTEGER	TOT%24<	NO OF SPELLING ERRORS THIS ESSAY

```
C     TSCONJ      OUT        INTEGER     TOT%26<      NO OF SUBORDINATING CONJ THIS ESS
C     TTITLE      OUT        INTEGER     TOT% 2<      TITLE THIS ESSAY
C     TTYPE       OUT        INTEGER     TOT%29<      NO OF DECLARATIVE TYPE A SENTENCE
C     TTYPE       OUT        INTEGER     TOT%30<      NO OF DECLARATIVE B S%NTENCES
C     TTYPE       OUT        INTEGER     TOT%31<      NO OF EXCLAMATORY SENTENCES
C     TTYPE       OUT        INTEGER     TOT%32<      NO OF QUESTIONS
C     TUNDER      OUT        INTEGER     TOT%15<      NO OF WORDS ITALICIZED THIS ESSAY
C     TWDSQ       OUT        INTEGER     TOT% 9<      SUM OF SQ OF WORDS IN EACH SENT
C     U           CHAR       REAL        ALPHA%21<    UBBBBB
C     V           CHAR       REAL        ALPHA%22<    VBBBBB
C     W           CHAR       REAL        ALPHA%23<    WBBBBB
C     X           CHAR       REAL        ALPHA%24<    XBBBBB
C     Y           CHAR       REAL        ALPHA%25<    YBBBBB
C     Z           CHAR       REAL        ALPHA%26<    ZBBBBB
C     COMMON IS AN AREA SHARED BY THIS PROGRAM AND ITS SUBPROGRAMS.
C
      COMMON/IN/BROKUP,TEXT,LENGTH                                      PEG I
C     BROKUP IS AN 80 CHARACTER CARD IMAGE.
      REAL BROKUP%80<                                                   PEG I
C     TEXT IS THE ASSEMBLED WORDS OF THE SENTENCE.
      DOUBLE PRECISION  TEXT%100<
C     COMMON SHARES THESE LISTS WITH OTHER PROGRAMS HAVING
C     COMMON/LISTS  IN THEIR TEXT.
      COMMON/LISTS/SUBCON,PREP,RELPRO,CONNEC,DALE,SPELLX,DECLAB,SVOPN   PEG I
C     TYPING ALL WORDS IN THESE ARRAYS AS DOUBLE PRECISION, I.E. THEY
C     CAN HAVE FROM 1-12 CHARACTERS EACH.
      DOUBLE PRECISION   SUBCON%20<,PREP%50<,RELPRO%10<,CONNEC%30<,DALE PEG I
     1%3000<,SPELLX%2000<,DECLAB%10<,SVOPN%150<                         PEG I
C     THE ARRAYS  PUNCT  AND  ALPHA  ARE SHARED WITH ALL PROGRAMS
C     HAVING THE    COMMON/CHAR/   STATEMENT.
      COMMON/CHAR/PUNCT,ALPHA                                           PEG I
C     TYPING THE ARRAYS  PUNCT  AND  ALPHA  AND THEIR CONTENTS AS REAL.
      REAL PUNCT%20<,BLANK,STAR,DECPT,COMMA,HYPHEN,APOSTR,OPAREN,CPAREN,PEG I
     1SLASH,PERIOD,COLON,SEMIC,EXCLAM,QUEST,DASH,ITALIC,ALPHA%26<,A,B,C,PEG I
     2D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z                     PEG I
C     THE EQUIVALENCE STATEMENT IS USED SO THAT THE CUMULATIVE ASPECTS
C     OF THE ARRAYS  PUNCT  AND  ALPHA  CAN BE INCORPORATED.
      EQUIVALENCE %PUNCT%1<,BLANK<,%PUNCT%2<,STAR<,%PUNCT%3<,DECPT<,%PUNPEG I
     1CT%4<,COMMA<,%PUNCT%5<,HYPHEN<,%PUNCT%6<,APOSTR<,%PUNCT%7<,OPAREN<PEG I
     2,%PUNCT%8<,CPAREN<,%PUNCT%9<,SLASH<,%PUNCT%10<,PERIOD<,%PUNCT%11<,PEG I
     3COLON<,%PUNCT%12<,SEMIC<,%PUNCT%13<,EXCLAM<,%PUNCT%14<,QUEST<,%PUNPEG I
     4CT%15<,DASH<,%PUNCT%16<,ITALIC<,%ALPHA%1<,A<,%ALPHA%2<,B<,%ALPHA%3PEG I
     5<,C<,%ALPHA%4<,D<,%ALPHA%5<,E<,%ALPHA%6<,F<,%ALPHA%7<,G<,%ALPHA%8<PEG I
     6,H<,%ALPHA%9<,I<,%ALPHA%10<,J<,%ALPHA%11<,K<,%ALPHA%12<,L<,%ALPHA%PEG I
     713<,M<,%ALPHA%14<,N<,%ALPHA%15<,O<,%ALPHA%16<,P<,%ALPHA%17<,Q<,%ALPEG I
     8PHA%18<,R<,%ALPHA%19<,S<,%ALPHA%20<,T<,%ALPHA%21<,U<,%ALPHA%22<,V<PEG I
     9,%ALPHA%23<,W<,%ALPHA%24<,X<,%ALPHA%25<,Y<,%ALPHA%26<,Z<          PEG I
C     THE ARRAYS  SUMS  AND  TOT  ARE SHARED WITH ALL PROGRAMS HAVING
C     THE   COMMON/OUT/   STATEMENT.
      COMMON/OUT/SUMS,TOT                                               PEG I
C     TYPING THE ARRAYS  SUMS  AND  TOT  AND THEIR CONTENTS AS INTEGER.
      INTEGER SUMS%100<,ID,TITLE,SENNUM,PARNUM,SUBVER,SUMLET,SSQLET,NUMWPEG I
     1DS,NWDSQ,NPAREN,NAPOS,NCOMMA,NPER,NPERCT,NUNDER,NDASH,NCOLO       PEG I
     2N,NSEMIC,NQUOTE,NEXCLA,NQUES,NPREP,NCONN,NSPELL,NRELPR,NSCONJ,NDALPEG I
     3E,ENDPCT,SENTYP%4<,TOT%100<,TID,TTITLE,NUMSEN,NUMPAR,SVOPEN,TOTLETPEG I
     4,TSQLET,TOTWDS,TWDSQ,TOTEND,TPAREN,TAPOS,TCOMMA,TPER,TPERCT,TUNDERPEG I
```

```
     5,TDASH,TCOLON,TSEMIC,TQUOTE,TEXCLA,TQUES,TPREP,TCONN,TSPELL,TRELPRPEG I
                                                                       PEG I
     6,TSCONJ,TDALE,TENDPT,TTYPE%4<
C     THE EQUIVALENCE STATEMENT IS USED SO THAT THE CUMULATIVE ASPECTS
C     OF THE ARRAY  SUMS  CAN BE INCORPORATED.
      EQUIVALENCE %SUMS%1<,ID<,%SUMS%2<,TITLE<,%SUMS%3<,SENNUM<,%SUMS%4<PEG I
     1,PARNUM<,%SUMS%5<,SUBVER<,%SUMS%6<,SUMLET<,%SUMS%7<,SSQLET<,%SUMS%PEG I
     28<,NUMWDS<,%SUMS%9<,NWDSQ <,              %SUMS%10<,NPAREN<,%SPEG I
     3UMS%11<,NAPOS <,%SUMS%12<,NCOMMA<,%SUMS%13<,NPER  <,%SUMS%14<,NPERPEG I
     4CT<,%SUMS%15<,NUNDER<,%SUMS%16<,NDASH<,%SUMS%17<,NCOLON<,%SUMS%18 PEG I
     5<,NSEMIC<,%SUMS%19<,NQUOTE<,%SUMS%20<,NEXCLA<,%SUMS%21<,NQUES <,%SPEG I
     6UMS%22<,NPREP<,%SUMS%23<,NCONN<,%SUMS%24<,NSPELL<,%SUMS%25<,NREL  PEG I
     7PR<,%SUMS%26<,NSCONJ<,%SUMS%27<,NDALE<,%SUMS%28<,ENDPCT<,%SUMS%29 PEG I
                                                                       PEG I
     8<,SENTYP<
C     THE EQUIVALENCE STATEMENT IS USED SO THAT THE CUMULATIVE ASPECTS
C     OF THE ARRAY  TOTS  CAN BE INCORPORATED.
      EQUIVALENCE%TOT%1<,TID<,%TOT%2<,TTITLE<,%TOT%3<,NUMSEN<,%TOT%4<,N PEG I
     1UMPAR<,%TOT%5<,SVOPEN<,%TOT%6<,TOTLET<,%TOT%7<,TSQLET<,%TOT%8<,TOTPEG I
     2WDS<,%TOT%9<,TWDSQ<,%TOT%10<,TPAREN<,%TOT%11<,TAP                 PEG I
     3OS<,%TOT%12<,TCOMMA<,%TOT%13<,TPER<,%TOT%14<,TPERCT<,%TOT%15<,TUNDPEG I
     4ER<,%TOT%16<,TDASH<,%TOT%17<,TCOLON<,%TOT%18<,TSEMIC<,%TOT%19<,TQUPEG I
     5OTE<,%TOT%20<,TEXCLA<,%TOT%21<,TQUES<,%TOT%22<,TPREP<,%TOT%23<,TCOPEG I
     6NN<,%TOT%24<,TSPELL<,%TOT%25<,TRELPR<,%TOT%26<,TSCONJ<,%TOT%27<,TDPEG I
                                                                       PEG I
     7ALE<,%TOT %28<,TENDPT<,%TOT%29<,TTYPE<
C     THESE ARE ADDED EQUIVALENCE STATEMENTS OTHERWISE THEY COULD HAVE
C     BEEN PLACED IN PRECEDING EQUIVALENCE STATEMENTS.
      EQUIVALENCE %SUMS%33<,NHYPH<,%TOT%33<,THYPH<
      EQUIVALENCE  %SUMS%34<,NSLASH<,%TOT%34<,TSLASH<
      INTEGER  NSLASH,TSLASH
      INTEGER  NHYPH,THYPH
C     MAKING NUNDER%NUMBER OF ITALICIZED WORDS THIS SENTENCE< AND NITAL
C     EQUIVALENT TO EACH OTHER.
      EQUIVALENCE  %NUNDER,NITAL<
C     TYPING AND DIMENSIONING WORD LENGTHS.
      INTEGER  LENGTH %100<
C     TYPING AND DIMENSIONING WORDS IN WORD LISTS.
                                                                       PEG I
      REAL RDTBL%10540<
C     THE EQUIVALENCE STATEMENT IS USED HERE TO SPECIFY WHAT PART OF THE
C     ARRAY %RDTBL< CORRESPONDS TO PARTICULAR WORD LISTS.
      EQUIVALENCE %RDTBL%1<,SUBCON<,%RDTBL%41<,PREP<,%RDTBL%141<,RELPRO<PEG I
     1,%RDTBL%161<,CONNEC<,%RDTBL%221<,DALE<,%RDTBL%6221<,SPELLX<,%RDTBLPEG I
                                                                       PEG I
     2%10221<,DECLAB<,%RDTBL%10241<,SVOPN<
C     DIMENSIONING AND TYPING  HLFTXT  WHICH REPRESENTS THE REAL
C     STORAGE OF THE CURRENT SENTENCE.
      REAL HLFTXT%200<
      EQUIVALENCE  %HLFTXT,TEXT<
C     VARIABLES SHARED BY OTHER PROGRAMS HAVING THE  COMMON/LOG/
C     STATEMENT.
      COMMON/LOG/SENTND,ESSEND
C     LOGICAL VARIABLES WHICH ARE SET TRUE OR FALSE DEPENDING UPON
C     LOGICAL TESTS.
      LOGICAL SENTND,ESSEND
C     MAKING ELEMENTS OF AN ARRAY EQUIVALENT TO CERTAIN PUNCTUATION
C     MARKS.
      EQUIVALENCE%PUNCT%17<,QUOTE<,%PUNCT%18<,PERCT<
C     TYPING QUOTES AND PER CENTS AS REAL VARIABLES.
      REAL QUOTE,PERCT
```

```
C       AN ARRAY%RELN< AND A VARIABLE SHARED BY OTHER PROGRAMS HAVING THE
C       COMMON/PSUM/   STATEMENT.
        COMMON/PSUM/  RELN,NEXT
C       TYPING THE ARRAY  RELN  AND THE VARIABLE  NEXT  AS INTEGERS.
        INTEGER  RELN%20<,NEXT
C      RELN %I< CONTAINS THESUMS SUBSCRIPT CORRESPONDING TO PUNCT%I<
C       AN ARRAY OF 10 WORDS SHARED BY OTHER PROGRAMS HAVING THE
C       COMMON/LIST2/   STATEMENT.
        COMMON/LIST2/ SWORD
C       TYPING THE ARRAY  SWORD  AS DOUBLE PRECISION, I.E. EACH WORD CAN
C       HAVE FROM 1-12 CHARACTERS.
        DOUBLE PRECISION SWORD%10<
C       TYPING THE ARRAY  SWRD  AS REAL.
        REAL  SWRD%20<
        EQUIVALENCE  %SWORD,SWRD<
C       A VARIABLE SHARED BY ALL PROGRAMS HAVING THE  COMMON/COUNT/
C       STATEMENT.
        COMMON/COUNT/ ICTR
C       TYPING THE VARIABLE  ICTR  AS AN INTEGER.
        INTEGER  ICTR
C       A LOGICAL VARIABLE SET TRUE OR FALSE DEPENDING UPON WORD BEING
C       ANALYZED IN THE SENTENCE.
        LOGICAL XX
C       A LOGICAL VARIABLE SET TRUE OR FALSE DEPENDING UPON WHETHER IT IS
C       THE START OF A NEW SENTENCE OR NOT.
        LOGICAL  START
C       INITIALIZING THE IMAGE COUNTER.
        ICTR#1
C       SEE MEMO FOR EXPLANATION%UNIVERSITY COMPUTER SYSTEM 7040-3<
        CALL FPTRAP
C       READ IN CARDS CONSISTING OF PUNCTUATION MARKS AND LETTERS OF THE
C       ALPHABET ACCORDING TO FORMAT STATEMENT 899 WHICH SPECIFIES
C       ARRANGEMENT.
        READ%5,899<  PUNCT,ALPHA
C       READ IN CARDS CONTAINING WORDS IN WORD TABLES.  THE SPLIT IN THE
C       ARRAY  RDTBL  IS A RESULT OF HAVING LESS THAN 2000 MISSPELLED
C       WORDS IN THE LIST.
        READ%5,900< %RDTBL%II<,II#1,7650<,  %RDTBL%II<,II#10221,10540<
C       READ IN CARDS CONSISTING OF 10 WORDS HAVING  S  ENDINGS.
        READ%5,908<  SWRD
C       2A6 SPECIFIES A DOUBLE PRECISION WORD CONSISTING OF 1-12
C       CHARACTERS.
908     FORMAT%2A6<
C       READ IN CARDS CONTAINING THE SUMS SUBSCRIPTS CORRESPONDING TO THE
C       PUNCTUATIONS.
        READ%5,905<  RELN
C       MEANS STORE THE LOGICAL CONSTANT TRUE IN START.
        START#.TRUE.
C       UNCONDITIONAL GO TO STATEMENT WHICH INTERRUPTS SEQUENTIAL
C       EXECUTION AND DIRECTS FLOW TO STATEMENT 15.
        GO TO 15
C
C   READ THE FIRST CARD OF THE NEXT ESSAY. IT CONTAINS THE ID NUMBER AND
C        INDICATION OF WHETHER OR NOT THERE IS A TITLE TO THIS ESSAY
C
C       READ IN FIRST CARD OF ESSAY CONTAINING IDENTIFICATION NUMBER AND
```

```
C       TITLE%IF ANY.<
2       READ%5,901<  IDENT,X
C       ASSIGN LOGICAL CONSTANT FALSE TO ESSEND%ESSAY END.<
        ESSEND#.FALSE.
C       A TEST TO DETERMINE WHETHER ALL THE ESSAYS ARE FINISHED.  IF A 999
C       CARD IS PRESENT THEN ALL THE ESSAYS HAVE BEEN PROCESSED.
        IF %IDENT.EQ.99999< GO TO 200
C       THIS STATEMENT SAYS THERE IS A TITLE BY ASSIGNING THE NUMBER
C       ONE TO TITLE.
        TITLE#1
C       THIS LOGICAL TEST THEN CHECKS TO SEE IF A TITLE EXISTS.  IF TRUE,
C       IT ASSIGNS THE PREDETERMINED VALUE ZERO.  IF FALSE, IT RETAINS
C       THE VALUE ASSIGNED IN THE PRECEDING STEP.
        IF%X.EQ.BLANK<  TITLE#0
C       REPLACE  ID  WITH THE IDENTIFICATION NUMBER.
        ID#IDENT
C       ASSIGN LOGICAL CONSTANT FALSE TO SENTND%SENTENCE END<
        SENTND#.FALSE.
C       INITIALIZE THE SEQUENCE NUMBER OF THE SENTENCE.
        SENNUM#0
C       INITIALIZE THE PARAGRAPH NUMBER.
        PARNUM # 1
C       READ THE 80 CHARACTER CARD IMAGE, ONE COLUMN PER CHARACTER.
        READ%5,902<  BROKUP
C       THIS ASSIGNED GO TO TRANSFERS CONTROL TO A CALL STATEMENT TO
C       DETERMINE SENTENCE ORGANIZATION.
        GO TO 100
C
C   READ IN EACH CHARACTER OF THE NEXT LINE OF THIS ESSAY
C   IF IT BEGINS A NEW ESSAY  OR IF IT IS A NEW PARAGRAPH FINISH THE
C    ANALYSES OF THE PREVIOUS SENTENCE AND ACCUMULATE TOTALS
C   OTHERWISE CONTINUE ASSEMBLING THIS CURRENT SENTENCE
C   UNTIL SENTENCE ENDING PUNCTUATION IS FOUND.
C
C       READ IN NEXT 80 CHARACTER CARD IMAGE.
1       READ%5,902<  BROKUP
C       IF FIRST OR SECOND CHARACTER NOT EQUAL TO A STAR OR BLANK
C       RESPECTIVELY GO TO 101.  %INCLUSIVE OR<  HAS VALUE TRUE IF EITHER
C       A OR B IS TRUE OR IF BOTH A AND B ARE TRUE.
        IF%BROKUP%1<.NE.STAR.OR.BROKUP%2<.NE.BLANK<  GO TO 101
C       IF ABOVE TEST FALSE THEN SET ESSAY END EQUAL TO TRUE.
        ESSEND#.TRUE.
C       IF ABOVE TEST FALSE SET SENTENCE END EQUAL TO TRUE.
        SENTND#.TRUE.
C       IF NEXT IS EQUAL TO ZERO SET SENTNECE END EQUAL TO FALSE.
        IF%NEXT.EQ.0<  SENTND#.FALSE.
C       IF NEXT IS EQUAL TO ZERO GO TO 15 AND INITIALIZE SUMS.
        IF%NEXT.EQ.0<  GO TO 15
C       GO TO  CALL SNTORG  FOR DETERMINATION OF SENTENCE ORGANIZATION.
        GO TO 100
101     DO 102 ICT#1,4
C       IF ANY OF THE FIRST FOUR IMAGES NOT BLANK GO TO   CALL SNTORG
C       FOR DETERMINATION OF SENTENCE ORGANIZATION.
102     IF%BROKUP%ICT<.NE.BLANK<  GO TO 100
C       INCREMENT PARAGRAPH NUMBER BY ONE.
        PARNUM#PARNUM&1
```

```
C       SET SENTENCE END EQUAL TO TRUE.
        SENTND#.TRUE.
        IF %NEXT.EQ.0<  SENTND#.FALSE.
C       THIS CALL STATEMENT CALLS THE SUBROUTINE SNTORG WHICH ANALYZES
C       EACH CHARACTER OF THE LINE OF THE ESSAY THAT WAS JUST READ IN.  IT
C       FINDS WORDS AND PUNCTUATION MARKS AND ASSEMBLES THE CHARACTERS
C       INTO SENTENCE COMPONENTS.  IT ALSO MAINTAINS A COUNT ON
C       SIGNIFICANT SENTENCE ELEMENTS.
100     CALL SNTORG
C       IF TRUE GO TO ONE AND READ IN NEXT CARD, IF FALSE CONTINUE WITH
C       NEXT STATEMENT.
        IF%.NOT.SENTND<  GO TO 1
C       MM  REPRESENTS THE NUMBER OF WORDS IN THE SENTENCE.
        MM#2*NEXT
C       WRITE THE SENTENCE ACCORDING TO THE FOLLOWING FORMAT.
        WRITE%6,906<  %HLFTXT%II<,IT#1,MM<
C       SPECIFIES TEN 12 CHARACTER WORDS PER LINE.
906     FORMAT%1H-20A6/%1H 20A6<<
C       INCREMENT THE NUMBER OF SENTENCES BY ONE.
        SENNUM#SENNUM&1
C       THIS SUBROUTINE TYPES EACH SENTENCE ACCORDING TO END PUNCTUATION
C       AND FIRST WORD TYPE.
        CALL TYPSEN
C       THIS SUBROUTINE DETERMINES TYPE OF SENTENCE OPENING.
        CALL SEEOPN
C
C          CHECK EACH WORD AGAINST THE LISTS OF SPECIAL WORDS
C                AND THE LIST OF COMMONLY MISSPELLED WORDS.
C
C       ICT IS INDEXED ACCORDING TO THE NUMBER OF WORDS IN THE SENTENCE.
        DO 10 ICT#1,NEXT
C       TESTING EACH WORD OF THE SENTENCE.  IF LENGTH%ICT< EQUAL TO 99
C       IT IS THE END OF THE SENTENCE.
        IF%LENGTH%ICT<.EQ.99<  GO TO 10
C
C       SET XX TRUE FOR FIRST OR SECOND WORD OF THE SENTENCE.
C       THIS WILL PREVENT IT FROM BEING CHECKED AGAINST THE LIST OF
C       RELATIVE PRONOUNS.
C
C       REPLACING  XX  BY ONE OR TWO FOR TEST IN SECOND STATEMENT BELOW.
        XX#ICT.EQ.1.OR.ICT.EQ.2
C       REFERS TO A SUBROUTINE WHICH DETERMINES THE TYPE OF WORD.
        CALL CHKLST%TEXT%ICT<,XX<
C       SEE COMMENT ABOVE WHICH STARTS WITH - SET XX TRUE FOR FIRST, ETC.
        IF%XX<  GO TO 10
C       REFERS TO A SUBROUTINE WHICH CHECKS FOR SPELLING.
        CALL SPELXX%TEXT%ICT<<
10      CONTINUE
C
C  PRINT THE RESUTLS OF THE ANALYSIS OF THIS SENTENCE AND, IF TI IS
C    THE LAST SENTENCE OF THE ESSAY, ALSO PRINT THE TOTAL RESULTS FOR
C       THE WHOLE ESSAY
C
C       WRITE SUMS FOR LINE ON TAPE 0.
        WRITE%0< %SUMS%II<,II#1,34<
C       PRINT SUMS FOR LINE.
```

```
      WRITE%6,903< %SUMS%II<,II#1,34<
C     ACCUMULATE TOTALS
      DO 11 ICT#1,4
11    TOT%ICT<#SUMS%ICT<
      DO 12 ICT#5,34
12    TOT%ICT<#TOT%ICT<&SUMS%ICT<
C     REPLACE SENTENCE END BY FALSE TO WORK WITH NEXT SENTENCE.
      SENTND#.FALSE.
C     INITIALIZING THE CUMULATIVE ASPECTS OF THE PROGRAM.  SUMS 1-5 DO
C     NOT REPRESENT CUMULATIVE ASPECTS OF THE PROGRAM.
15    DO 13 ICT#5,100
13    SUMS%ICT<#0
C     INITIALIZING NEXT FOR NEXT SENTENCE.
      NEXT#0
C     TEST TO DETERMINE IF BEGINNING OF ANOTHER LINE.
      IF%START<  GO TO 16
C     IF TRUE GO TO SNTORG.  IF FALSE, THEN WRITE TOTALS FOR ESSAY.
      IF%.NOT.ESSEND.AND.ICTR.GT.80< GO TO 1
      IF%.NOT. ESSEND<  GO TO 100
C     WRITING ACCUMULATED TOTALS FOR AN ESSAY ON TAPE UNIT 0.
      WRITE%0< %TOT%II<,II#1,34<
C     WRITING ACCUMULATED TOTALS FOR AN ESSAY ON TAPE UNIT 4.
      WRITE%4< %TOT%II<,II#1,34<
C     PRINTED OUTPUT FOR ACCUMULATED TOTALS FOR AN ESSAY.
      WRITE%6,903< %TOT%II<,II#1,34<
C     TYPING THE VARIABLE X11%31 VARIABLES< FOR PUNCHED AND PRINTED
C     OUTPUT.
      REAL X11%31<
C     IDENTIFICATION FOR PRINTED AND PUNCHED OUTPUT FOR ESSAY.
      I1#1
      I2#2
C     TOT%3< IS NUMBER OF SENTENCES THIS ESSAY.
      SEN#TOT%3<
C     TOT%8< IS NUMBER OF WORDS THIS ESSAY.
      WDS#TOT%8<
C     TOT%2< REPRESENTS TITLE THIS ESSAY.
      X11%1<#TOT%2<
C     A RATIO SCALE FOR ONE OF THE VARIABLES.
      X11%2<#%WDS/SEN<*10.
C     TOT%4< IS NUMBER OF PARAGRAPHS THIS ESSAY.
      X11%3<#TOT%4<
C     FLOAT TOT%5< MEANS MAKE THE INTEGER TOT%5< REAL AND IT REPRESENTS
C     THE NUMBER OF S-V SENTENCE OPENINGS.
      X11%4<#%FLOAT%TOT%5<</SEN<*100.
      X11%5<#WDS
C     THE DO 511 LOOP ESTABLISHES THE NUMBER OF PUNCTUATION MARKS IN
C     THE ESSAY ACCORDING TO A PREDETERMINED SCALE.
      DO 511 II#6,14
511   X11%II<#%FLOAT%TOT%II&4<</WDS<*1000.
C     THE DO 512 LOOP ESTABLISHES THE NUMBER OF QUOTES, NUMBER OF
C     QUESTION MARKS, AND NUMBER OF EXCLAMATION POINTS IN THE ESSAY
C     ACCORDING TO A PREDETERMINED SCALE.
      DO 512 II#15,17
512   X11%II<#%FLOAT%TOT%II&4<</SEN<*1000.
C     THE DO 513 LOOP ESTABLISHES THE NUMBER OF PREPOSITIONS,
C     CONNECTIVES, SPELLING ERRORS, RELATIVE PRONOUNS, SUBORDINATING
```

```
C     CONJUNCTIONS, AND DALF WORDS FOR THE ESSAY.
      DO 513 II#18,23
513   X11%II<#%FLOAT%TOT%II&4<</WDS<*100.
C     THE DO 514 LOOP ESTABLISHES THE NUMBER OF SENTENCES WITH NO
C     ENDING PUNCTUATION, DECLARATIVE TYPE A SENTENCES, AND DECLARATIVE
C     TYPE B SENTENCES IN THIS ESSAY.
      DO 514 II#24,26
514   X11%II<#%FLOAT%TOT%II&4<</SEN<*100.
C     SCALE FOR SUM OF HYPHENS IN ESSAY.
      X11%27<#%FLOAT%TOT% 33<</WDS<*1000.
C     SCALE FOR SUM OF SLASHES IN ESSAY.
      X11%28<#%FLOAT%TOT%34<</WDS<*1000.
C     SCALE FOR SUM OF LETTERS THIS ESSAY.
      X11%29<#%FLOAT%TOT%6<</WDS<*100.
C     %SQRT OF SUM OF SQUARED LETTERS %TOT%7<< IN ESSAY DIVIDED BY THE
C     NUMBER OF WORDS MINUS THE SUM OF LETTERS DIVIDED BY 100
C     SQUARED . THIS QUANTITY IS THEN MULTIPLIED BY A 100.
      X11%30<#SQRT%FLOAT%TOT%7<</WDS-%X11%29</100.<**2<*100.
C     TEN TIMES THE SQUARE ROOT OF THE QUANTITY SUM OF SQUARE OF WORDS
C     IN EACH SENTENCE DIVIDED BY THE NUMBER OF SENTENCES MINUS THE
C     NUMBER OF WORDS DIVIDED BY THE NUMBER OF SENTENCES TIMES 10
C     SQUARE.
      X11%31<#SQRT%FLOAT%TOT%9<</SEN-%X11%2</10.<**2<*10.
C  PUNCH OUTPUT CARD
C     PRINTED OUTPUT FOR AN ESSAY.
      WRITE%6,965< ID, I1,%X11%II<,II#1,9<,%X11%II<,II#11,16<,ID,I2,
     1%X11%II<,II#17,31<
C     PUNCHED CARD OUTPUT FOR AN ESSAY.
      WRITE%7,904< ID, I1,%X11%II<,II#1,9<,%X11%II<,II#11,16<,ID,I2,
     1%X11%II<,II#17,31<
C     REPLACE START BY FALSE SO THAT THE PROGRAM WILL NOT INTERPRET
C     START AS THE BEGINNING OF ANOTHER LINE.
16    START#.FALSE.
C     INITIALIZE IDENTIFICATION NUMBER, TITLE, SEQUENCE NUMBER OF
C     SENTENCE, AND SEQUENCE NUMBER OF PARAGRAPH.
      DO 14  ICT#1,4
14    SUMS%ICT<#0
C     THE DO 17 LOOP INITIALIZES THE CUMULATIVE ASPECTS OF THE
C     PROGRAM FOR EACH ESSAY.
      DO 17 ICT#5,34
17    TOT%ICT<#0
C     BEGIN ANALYZING THE ESSAY BY GOING TO 2.
      GO TO 2
C     SET IDENTIFICATION NUMBER EQUAL TO TOT%1<.
200   TOT%1<#IDENT
C     WRITE THE TOTALS FOR ESSAY ON TAPE 0.
      WRITE%0< %TOT%II<,II#1,34<
C     WRITE THE TOTALS FOR THE ESSAY ON TAPE 4.
      WRITE%4< %TOT%II<,II#1,34<
C     LAST ENTRY MADE ON TAPE 0.
      END FILE 0
C     LAST ENTRY MADE ON TAPE 4.
      END FILE 4
C     REWIND TAPE 0 %END OF JOB<
      REWIND  0
C     REWIND TAPE 4 %END OF JOB<
```

APPENDIX A (Continued)

```
      REWIND 4
      RETURN
C     SPECIFIES 6 CHARACTER WORDS FOR PUNCT AND ALPHA.
899   FORMAT%A6<
C     SPECIFIES SIX 12 CHARACTER WORDS PER LINE FOR WORD LISTS.
900   FORMAT%12A6<
C     SKIP 7 SPACES, IDENTIFICATION NUMBER, SKIP 2 SPACES, TITLE
901   FORMAT%I5,9X,A1<
C     SPECIFIES 80 SUCCESSIVE FIELDS OF ONE CHARACTER EACH.
902   FORMAT%80A1<
C     SPECIFIES PRINT OUT FOR SUMS AND TOTS.
903   FORMAT%1H0I5,I2,3I3,2I5,2I5,17I3,I4,5I2,2I3<
C     SPECIFIES PUNCHED CARD OUTPUT FOR AN ESSAY%TOTALS<
904   FORMAT%I5,I1,F4.0,F5.0,F4.0,12F5.0<
C     SPECIFIES PRINT OUT FOR AN ESSAY%TOTALS<
965   FORMAT%1XI5,I1,15F5.0<
C     SPECIFIES 20 INTEGERS%1 OR 2 DIGITS< FOR RELN.
905   FORMAT%20I2<
      END
```

```
$IBFTC SNTRG
      SUBROUTINE   SNTORG                                             PEG I
      COMMON/IN/BROKUP,TEXT,LENGTH
      REAL BROKUP%80<,BRKUP%3<
      DOUBLE PRECISION  TEXT%100<
      COMMON/LISTS/SUBCON,PREP,RELPRO,CONNEC,DALE,SPELLX,DECLAB,SVOPN   PEG I
      DOUBLE PRECISION   SUBCON%20<,PREP%50<,RELPRO%10<,CONNEC%30<,DALE PEG I
     1%3000<,SPELLX%2000<,DECLAB%10<,SVOPN%150<                         PEG I
      COMMON/CHAR/PUNCT,ALPHA                                           PEG I
      REAL PUNCT%20<,BLANK,STAR,DECPT,COMMA,HYPHEN,APOSTR,OPAREN,CPAREN,PEG I
     1SLASH,PERIOD,COLON,SEMIC,EXCLAM,QUEST,DASH,ITALIC,ALPHA%26<,A,B,C,PEG I
     2D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z                     PEG I
      EQUIVALENCE %PUNCT%1<,BLANK<,%PUNCT%2<,STAR<,%PUNCT%3<,DECPT<,%PUNPEG I
     1CT%4<,COMMA<,%PUNCT%5<,HYPHEN<,%PUNCT%6<,APOSTR<,%PUNCT%7<,OPAREN<PEG I
     2,%PUNCT%8<,CPAREN<,%PUNCT%9<,SLASH<,%PUNCT%10<,PERIOD<,%PUNCT%11<,PEG I
     3COLON<,%PUNCT%12<,SEMIC<,%PUNCT%13<,EXCLAM<,%PUNCT%14<,QUEST<,%PUNPEG I
     4CT%15<,DASH<,%PUNCT%16<,ITALIC<,%ALPHA%1<,A<,%ALPHA%2<,B<,%ALPHA%3PEG I
     5<,C<,%ALPHA%4<,D<,%ALPHA%5<,E<,%ALPHA%6<,F<,%ALPHA%7<,G<,%ALPHA%8<PEG I
     6,H<,%ALPHA%9<,I<,%ALPHA%10<,J<,%ALPHA%11<,K<,%ALPHA%12<,L<,%ALPHA%PEG I
     713<,M<,%ALPHA%14<,N<,%ALPHA%15<,O<,%ALPHA%16<,P<,%ALPHA%17<,Q<,%ALPEG I
     8PHA%18<,R<,%ALPHA%19<,S<,%ALPHA%20<,T<,%ALPHA%21<,U<,%ALPHA%22<,V<PEG I
     9,%ALPHA%23<,W<,%ALPHA%24<,X<,%ALPHA%25<,Y<,%ALPHA%26<,Z<          PEG I
      COMMON/OUT/SUMS,TOT                                               PEG I
      INTEGER SUMS%100<,ID,TITLE,SENNUM,PARNUM,SUBVER,SUMLET,SSQLET,NUMWPEG I
     1DS,NWDSQ,NPAREN,NAPOS,NCOMMA,NPER,NPERCT,NUNDER,NDASH,NCOLO      PEG I
     2N,NSEMIC,NQUOTE,NEXCLA,NQUES,NPREP,NCONN,NSPELL,NRELPR,NSCONJ,NDALPEG I
     3E,ENDPCT,SENTYP%4<,TOT%100<,TID,TTITLE,NUMSEN,NUMPAR,SVOPEN,TOTLETPEG I
     4,TSQLET,TOTWDS,TWDSQ,TOTFND,TPAREN,TAPOS,TCOMMA,TPER,TPERCT,TUNDERPEG I
     5,TDASH,TCOLON,TSEMIC,TQUOTE,TEXCLA,TQUES,TPREP,TCONN,TSPELL,TRELPRPEG I
     6,TSCONJ,TDALE,TENDPT,TTYPE%4<                                     PEG I
      EQUIVALENCE %SUMS%1<,ID<,%SUMS%2<,TITLE<,%SUMS%3<,SENNUM<,%SUMS%4<PEG I
     1,PARNUM<,%SUMS%5<,SUBVER<,%SUMS%6<,SUMLET<,%SUMS%7<,SSQLET<,%SUMS%PEG I
     28<,NUMWDS<,%SUMS%9<,NWDSQ <,             %SUMS%10<,NPAREN<,%SPEG I
     3UMS%11<,NAPOS <,%SUMS%12<,NCOMMA<,%SUMS%13<,NPER  <,%SUMS%14<,NPERPEG I
     4CT<,%SUMS%15<,NUNDER<,%SUMS%16<,NDASH<,%SUMS%17<,NCOLON<,%SUMS%18 PEG I
     5<,NSEMIC<,%SUMS%19<,NQUOTE<,%SUMS%20<,NEXCLA<,%SUMS%21<,NQUES <,%SPEG I
     6UMS%22<,NPREP<,%SUMS%23<,NCONN<,%SUMS%24<,NSPELL<,%SUMS%25<,NREL  PEG I
```

```
     7PR<,%SUMS%26<,NSCONJ<,%SUMS%27<,NDALE<,%SUMS%28<,ENDPCT<,%SUMS%29 PEG I
     8<,SENTYP<                                                         PEG I
      EQUIVALENCE%TOT%1<,TID<,%TOT%2<,TTITLE<,%TOT%3<,NUMSEN<,%TOT%4<,N PEG I
     1UMPAR<,%TOT%5<,SVOPEN<,%TOT%6<,TOTLET<,%TOT%7<,TSQLET<,%TOT%8<,TOTPEG I
     2WDS<,%TOT%9<,TWDSQ<,%TOT%10<,TPAREN<,%TOT%11<,TAP                 PEG I
     3OS<,%TOT%12<,TCOMMA<,%TOT%13<,TPER<,%TOT%14<,TPERCT<,%TOT%15<,TUNDPEG I
     4ER<,%TOT%16<,TDASH<,%TOT%17<,TCOLON<,%TOT%18<,TSEMIC<,%TOT%19<,TQUPEG I
     5OTE<,%TOT%20<,TEXCLA<,%TOT%21<,TQUES<,%TOT%22<,TPREP<,%TOT%23<,TCOPEG I
     6NN<,%TOT%24<,TSPELL<,%TOT%25<,TRELPR<,%TOT%26<,TSCONJ<,%TOT%27<,TDPEG I
     7ALE<,%TOT %28<,TENDPT<,%TOT%29<,TTYPE<                            PEG I
      EQUIVALENCE %SUMS%33<,NHYPH<,%TOT%33<,THYPH<
      EQUIVALENCE  %SUMS%34<,NSLASH<,%TOT%34<,TSLASH<
      INTEGER  NSLASH,TSLASH
      INTEGER  NHYPH,THYPH
      EQUIVALENCE  %NUNDER,NITAL<
      INTEGER  LENGTH %100<
      REAL RDTBL%10540<                                                 PEG I
      EQUIVALENCE %RDTBL%1<,SUBCON<,%RDTBL%41<,PREP<,%RDTBL%141<,RELPRO<PEG I
     1,%RDTBL%161<,CONNEC<,%RDTBL%221<,DALE<,%RDTBL%6221<,SPELLX<,%RDTBLPEG I
     2%10221<,DECLAB<,%RDTBL%10241<,SVOPN<                              PEG I
      REAL HLFTXT%200<
      EQUIVALENCE  %HLFTXT,TEXT<
      COMMON/LOG/SENTND,ESSEND
      LOGICAL SENTND,ESSEND
      EQUIVALENCE%PUNCT%17<,QUOTE<,%PUNCT%18<,PERCT<
      REAL QUOTE,PERCT
      COMMON/PSUM/  RELN,NEXT
      INTEGER  RELN%20<,NEXT
C    RELN %I< CONTAINS THESUMS SUBSCRIPT CORRESPONDING TO PUNCT%I<
      COMMON/LIST2/ SWORD
      DOUBLE PRECISION SWORD%10<
      REAL  SWRD%20<
      EQUIVALENCE  %SWORD,SWRD<
C
C   SNTORG  ANALYSES EACH CHARACTER OF THE LINE OF THE ESSAY THAT
C     WAS JUST READ IN.  IT FINDS WORDS AND PUNCTUATION AND
C   ASSEMBLES THE CHARACTERS INTO SENTENCE COMPONENTS
C   THIS PART ALSO MAINTAINS COUNTS ON SIGNIFICANT SENTENCE ELEMENTS
C
C     AN AREA SHARED BY THIS SUBROUTINE AND THE MAIN PROGRAM.
      COMMON/COUNT/ ICTR
C     TYPING IMAGE COUNTER AS AN INTEGER.
      INTEGER  ICTR
C     TYPING THE VARIABLES ONE, TWO, AND THREE AS LOGICAL.
      LOGICAL  ONE,TWO,THREE
C     TYPING CT AND IRELN AS INTEGERS.
      INTEGER CT,IRELN
C     TYPING TEMPA AS A DOUBLE PRECISION WORD.
      DOUBLE PRECISION  TEMPA
C     TYPING THE ARRAY TEMPB AS REAL.
      REAL  TEMPB%2<
C     MAKING TEMPA AND TEMPB EQUIVALENT.
      EQUIVALENCE  %TEMPB,TEMPA<
C     AN AREA SHARED BY THIS SUBROUTINE AND SUBROUTINE TYPSEN%TYPE OF
C     SENTENCE<.
      COMMON/ENDS/ ALSPER,ALSEXC,ALSQUS
```

```
C       TYPING THE ARRAYS AS REAL.
        REAL  ALSPER%2<,ALSEXC%2<,ALSQUS%2<
C       THIS CALL STATEMENT RESULTS IN SETTING ALSPER EQUAL TO A
C       PUNCTUATION MARK %PERIOD< WITHIN QUOTES.
        CALL DATA %ALSPER,6H.*     <
C       THIS CALL STATEMENT RESULTS IN SETTING ALSEXC EQUAL TO A
C       PUNCTUATION MARK%EXCLAMATION POINT< WITHIN QUOTES.
        CALL DATA %ALSEXC,6H.X*    <
C       THIS CALL STATEMENT RESULTS IN SETTING ALSQUS EQUAL TO A
C       PUNCTUATION MARK%QUESTION MARK< WITHIN QUOTES.
        CALL DATA %ALSQUS,6H.Q*    <
C       A LOGICAL TEST TO DETERMINE IF SENTND IS TRUE.  IF SO, RETURN TO
C       THE MAIN PROGRAM.  IF NOT, CONTINUE.
C       INITIALIZE LETTER COUNTER.
        LCTR#0
C       A LOGICAL TEST TO DETERMINE IF IMAGE COUNTER GREATER THAN OR EQUAL
C       TO 81.  IF SO, REPLACE ICTR BY ONE AND CONTINUE.  IF NOT, CONTINUE
        IF%ICTR.GE.81<  ICTR#1
        IF%SENTND<  RETURN
C       SAME AS PRECEDING STATEMENT EXCEPT RETURN IF TRUE, I.E. THE CARD
C       IMAGES HAVE ALL BEEN ANALYZED.
1       IF%ICTR.GE.81.AND.LCTR.NE.0< GO TO 200
        IF%ICTR.GE.81< RETURN
C       SET LOGICAL CONSTANT TRUE EQUAL TO ONE.
        ONE#.TRUE.
C       SET LOGICAL CONSTANT TRUE EQUAL TO TWO.
        TWO#.TRUE.
C       SET LOGICAL CONSTANT TRUE EQUAL TO THREE.
        THREE#.TRUE.
C       THE DO 2 LOOP IS ESTABLISHED TO FIND A PUNCTUATION MARK.  IF SO,
C       BREAK OUT OF THE LOOP.  IF NOT, CONTINUE.
        DO 2 CT#2,9
C       A LOGICAL TEST TO DETERMINE IF IMAGES 2-9 ARE PUNCTUATION MARKS
C       %ASTERISK%2<,DECIMAL POINT%3<,COMMA%4<,HYPHEN%5<,APOSTROPHE%6<,
C       OPEN PARENTHESIS%7<,CLOSED PARENTHESIS%8<,SLASH%9<<.  IF SO, GO
C       TO THREE.
2       IF%BROKUP%ICTR<.EQ.PUNCT%CT<<  GO TO 3
C       A LOGICAL TEST TO DETERMINE IF THERE IS AN IMAGE.  IF SO, GO TO
C       300 TO WORK WITH THE IMAGE.  IF NOT, CONTINUE.
        IF%BROKUP%ICTR<.NE.BLANK<  GO TO 300
C       A LOGICAL TEST TO DETERMINE IF THE LETTER COUNTER IS NOT EQUAL TO
C       ZERO.  IF SO, GO TO 200.  IF NOT, CONTINUE.
        IF%LCTR.NE.0<  GO TO 200
C       IF ABOVE THREE TESTS FALSE GO TO 400.
        GO TO 400
C       PLACE THE PUNCTUATION MARK IN A WORD AND THEN COMPARE WITH
C       PREESTABLISHED WORD PUNCTUATION MARKS.
3       BRKUP%1<#BROKUP%ICTR<
        BRKUP%2<#BROKUP%ICTR&1<
        BRKUP%3<#BROKUP%ICTR&2<
        IF%ICTR.GT.79< BRKUP%2<#BLANK
        IF%ICTR.GT.78< BRKUP%3<#BLANK
        CALL PACK%BRKUP%1<,TEMPA,3<
C       IF ANY OF THESE ARE TRUE OR ALL ARE TRUE AND LETTER COUNTER IS
C       EQUAL TO ZERO GO TO 100 FOR SUMMATIONS BECAUSE IT IS THE END OF
C       A SENTENCE.
```

```
      IF%%%TEMPB%1<.EQ.PERIOD<.OR.%TEMPB%1<.EQ.EXCLAM<.OR.%TEMPB%1<.EQ.Q
     1UEST<<.AND.LCTR.EQ.0<  GO TO 100
C     SAME AS COMMENT ABOVE.
      IF%%%TEMPB%1<.EQ.ALSPER<.OR.%TEMPB%1<.EQ.ALSEXC<.OR.%TEMPB%1<.EQ.
     1ALSQUS<<.AND.LCTR.EQ.0<   GO TO 100
C     REPLACE THREE BY FALSE.
      THREE#.FALSE.
C     IF LETTER COUNTER NOT EQUAL TO ZERO GO TO 4 FOR FURTHER CHECKING.
      IF%LCTR.NE.0<  GO TO 4
C     IF TEMPB%1< EQUAL TO AN ITALIC MARK GO TO 100 FOR FURTHER
C     PROCESSING.
      IF%TEMPB%1<.EQ.ITALIC<  GO TO 100
      CALL PACK%BRKUP%1<,TEMPA,2<
C     REPLACE TWO BY FALSE.
      TWO #.FALSE.
C     THE DO 5 LOOP SETS UP THE APPROPRIATE INDEX FOR STATEMENTS
C     FOLLOWING STATEMENT 100 AND ALSO DETERMINES IF TEMPB%1< IS EQUAL
C     TO A PERIOD, COLON, SEMICOLON, EXCLAMATION, QUESTION, DASH,
C     ITALIC, QUOTE, OR A PER CENT.
      DO 5 CT#10,18
      IRELN#RELN%CT<
5     IF%TEMPB%1<.EQ.PUNCT%CT<<  GO TO 100
C     REPLACE TWO BY TRUE.
      TWO #.TRUE.
C     REPLACE ONE BY FALSE.
      ONE#.FALSE.
C     AN ASSIGNED GO TO STATEMENT WHICH ELIMINATES FOLLOWING TEST.
      GO TO 100
C     IF IMAGE IS EQUAL TO AN APOSTROPHE GO TO 500.
4     IF%BROKUP%ICTR<.EQ.APOSTR<  GO TO 500
      GO TO 200
C     ASSIGN THE MINIMUM VALUE OF THE TWO ARGUMENTS EQUAL TO NEXT.
100   NEXT#MINO%NEXT&1,100<
C     REPLACE LENGTH BY 99 WHICH IS THE END OF THE SENTENCE.
      LENGTH%NEXT<#99
C     IF TWO IS TRUE GO TO 101 AND BEGIN SUMMING PUNCTUATION MARKS.
      IF%TWO<  GO TO 101
C     SUM THE APPROPRIATE PUNCTUATION MARKS AS INDICATED BY THE
C     SUBSCRIPT IRELN.
      SUMS%IRELN<#SUMS%IRELN<&1
C     AN ASSIGNED GO TO STATEMENT WHICH ELIMINATES THE FOLLOWING 9
C     TESTS.
      GO TO 102
C     IF ONE AND NOT THREE ARE TRUE INCREASE THE NUMBER OF ITALICIZED
C     WORDS BY ONE.
101   IF%ONE.AND.%.NOT.THREE<<  NITAL#NITAL&1
C     IF NOT ONE OR THREE TRUE GO TO 107.
      IF%.NOT.%ONE.OR.THREE<<  GO TO 107
C     SET END OF SENTENCE PUNCTUATION EQUAL TO ONE.
      ENDPCT#1
C     IF TRUE INCREASE THE NUMBER OF PERIODS BY ONE.
      IF%TEMPB%1<.EQ.PERIOD<  NPER#NPER&1
C     IF TRUE INCREASE THE NUMBER OF PERIODS BY ONE.
      IF%TEMPB%1<.EQ.ALSPER<  NPER#NPER&1
C     IF TRUE INCREASE THE NUMBER OF EXCLAMATIONS BY ONE.
      IF%TEMPB%1<.EQ.EXCLAM<  NEXCLA#NEXCLA&1
```

```
C       IF TRUE INCREASE THE NUMBER OF EXCLAMATIONS BY ONE.
        IF%TEMPB%1<.EQ.ALSEXC<  NEXCLA#NEXCLA&1
C       IF TRUE INCREASE THE NUMBER OF QUESTION MARKS BY ONE.
        IF%TEMPB%1<.EQ.QUEST< NQUES#NQUES&1
C       IF TRUE INCREASE THE NUMBER OF QUESTION MARKS BY ONE.
        IF%TEMPB%1<.EQ.ALSQUS< NQUES#NQUES&1
C       ASSIGNED GO TO STATEMENT.
        GO TO 102
C       THE DO 109 LOOP DETERMINES IF ANY OF THE IMAGES %2-9< ARE
C       PUNCTUATION MARKS.  IF SO, REPLACE IRELN BY RELN%CT<.
107     DO 109  CT#2,9
109     IF%BROKUP%ICTR<.EQ.PUNCT%CT<<  IRELN#RELN%CT<
C       INCREMENT SUMS BY ONE FOR APPROPRIATE PUNCTUATION MARK.
        SUMS%IRELN<#SUMS%IRELN<&1
C       IF ONE IS TRUE INCREMENT LETTER COUNTER BY TWO.
102     IF%ONE<  LCTR#LCTR&2
C       IF TWO IS TRUE INCREMENT LETTER COUNTER BY ONE.
        IF%TWO<  LCTR#LCTR&1
        CALL PACK%BRKUP%1<,TEXT%NEXT<,LCTR<
C       REPLACE IMAGE COUNTER BY VALUE IN IMAGE COUNTER PLUS VALUE IN
C       LETTER COUNTER.
        ICTR#ICTR&LCTR
C       INITIALIZE LETTER COUNTER.
        LCTR#0
C       IF NOT THREE TRUE GO TO ONE, OTHERWISE CONTINUE.
        IF%.NOT.THREE<  GO TO 1
C       DECREMENT IMAGE COUNTER BY ONE.
        ICTR#ICTR-1
C       SET SENTENCE END EQUAL TO TRUE.
        SENTND#.TRUE.
C       REPLACE NUMBER OF WORDS SQUARED BY NUMBER OF WORDS TIMES
C       NUMBER OF WORDS.
        NWDSQ#NUMWDS*NUMWDS
C       RETURN TO ORIGINAL PROGRAM.
        RETURN
C       SET NEXT EQUAL TO THE MINIMUM VALUE OF THE TWO ARGUMENTS.
200     NEXT#MINO%NEXT&1,100<
C       LENGTH OF THE PARTICULAR WORD IS SET EQUAL TO THE CONSTANT IN
C       LETTER COUNTER%LCTR<.
        LENGTH%NEXT<#LCTR
C       REPLACE IBACK WITH IMAGE COUNTER VALUE MINUS LETTER COUNTER
C       VALUE.
        IBACK#ICTR-LCTR
C       PACKING EACH WORD FOR ANALYSIS.
        CALL PACK%BROKUP%IBACK<,TEXT%NEXT<,LCTR<
C       REPLACE NUMBER OF WORDS BY NUMBER OF WORDS PLUS ONE %ACCUMULATING
C       WORDS IN SENTENCE<.
        NUMWDS#NUMWDS&1
C       REPLACE SUM OF SQUARED LETTERS BY WHAT IS IN SUM OF SQUARED
C       LETTERS PLUS LETTER COUNTER VALUE SQUARED.
        SSQLET#SSQLET&LCTR*LCTR
C       REPLACE LETTER COUNTER WITH 0 AND CONTINUE WITH NEXT WORD.
        LCTR#0
C       CONTINUE WITH NEXT IMAGE.
        GO TO 1
C       SUMMING THE LETTERS IN THE SENTENCE.
```

```
300   SUMLET#SUMLET&1
C     FINDING THE LENGTH OF A WORD.
      LCTR#LCTR&1
C     SUMMING THE IMAGES ON A CARD TO DETERMINE WHEN ALL IMAGES HAVE
C     BEEN PROCESSED.
      ICTR#ICTR&1
C     CONTINUE WITH NEXT IMAGE.
      GO TO 1
C     SUMMING THE IMAGES ON A CARD TO DETERMINE WHEN ALL IMAGES HAVE
C     BEEN PROCESSED.
400   ICTR#ICTR&1
C     CONTINUE WITH NEXT IMAGE.
      GO TO 1
C     INCREASE THE NUMBER OF APOSTROPHES BY ONE.
500   NAPOS#NAPOS&1
C     INCREMENT LETTER COUNTER BY ONE.
      LCTR#LCTR&1
C     INCREMENT IMAGE COUNTER BY ONE.
      ICTR#ICTR&1
C     CONTINUE WITH PROGRAM.
      GO TO 1
      END
$IBFTC SP
      SUBROUTINE  SPELXX  %WORD<
      COMMON/IN/BROKUP,TEXT,LENGTH                                      PEG I
      REAL BROKUP%80<                                                   PEG I
      DOUBLE PRECISION  TEXT%100<
      COMMON/LISTS/SUBCON,PREP,RELPRO,CONNEC,DALE,SPELLX,DECLAB,SVOPN   PEG I
      DOUBLE PRECISION   SUBCON%20<,PREP%50<,RELPRO%10<,CONNEC%30<,DALE PEG I
     1%3000<,SPELLX%2000<,DECLAB%10<,SVOPN%150<                         PEG I
      COMMON/CHAR/PUNCT,ALPHA                                           PEG I
      REAL PUNCT%20<,BLANK,STAR,DECPT,COMMA,HYPHEN,APOSTR,OPAREN,CPAREN,PEG I
     1SLASH,PERIOD,COLON,SEMIC,EXCLAM,QUEST,DASH,ITALIC,ALPHA%26<,A,B,C,PEG I
     2D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z                     PEG I
      EQUIVALENCE %PUNCT%1<,BLANK<,%PUNCT%2<,STAR<,%PUNCT%3<,DECPT<,%PUNPEG I
     1CT%4<,COMMA<,%PUNCT%5<,HYPHEN<,%PUNCT%6<,APOSTR<,%PUNCT%7<,OPAREN<PEG I
     2,%PUNCT%8<,CPAREN<,%PUNCT%9<,SLASH<,%PUNCT%10<,PERIOD<,%PUNCT%11<,PEG I
     3COLON<,%PUNCT%12<,SEMIC<,%PUNCT%13<,EXCLAM<,%PUNCT%14<,QUEST<,%PUNPEG I
     4CT%15<,DASH<,%PUNCT%16<,ITALIC<,%ALPHA%1<,A<,%ALPHA%2<,B<,%ALPHA%3PEG I
     5<,C<,%ALPHA%4<,D<,%ALPHA%5<,E<,%ALPHA%6<,F<,%ALPHA%7<,G<,%ALPHA%8<PEG I
     6,H<,%ALPHA%9<,I<,%ALPHA%10<,J<,%ALPHA%11<,K<,%ALPHA%12<,L<,%ALPHA%PEG I
     713<,M<,%ALPHA%14<,N<,%ALPHA%15<,O<,%ALPHA%16<,P<,%ALPHA%17<,Q<,%ALPEG I
     8PHA%18<,R<,%ALPHA%19<,S<,%ALPHA%20<,T<,%ALPHA%21<,U<,%ALPHA%22<,V<PEG I
     9,%ALPHA%23<,W<,%ALPHA%24<,X<,%ALPHA%25<,Y<,%ALPHA%26<,Z<          PEG I
      COMMON/OUT/SUMS,TOT                                               PEG I
      INTEGER SUMS %100<,ID,TITLE,SENNUM,PARNUM,SUBVER,SUMLET,SSQLET,NUMWPEG I
     1DS,NWDSQ,NPAREN,NAPOS,NCOMMA,NPER,NPERCT,NUNDER,NDASH,NCOLO       PEG I
     2N,NSEMIC,NQUOTE,NEXCLA,NQUES,NPREP,NCONN,NSPELL,NRELPR,NSCONJ,NDALPEG I
     3E,ENDPCT,SENTYP%4<,TOT%100<,TID,TTITLE,NUMSEN,NUMPAR,SVOPEN,TOTLETPEG I
     4,TSQLET,TOTWDS,TWDSQ,TOTFND,TPAREN,TAPOS,TCOMMA,TPER,TPERCT,TUNDERPEG I
     5,TDASH,TCOLON,TSEMIC,TQUOTE,TEXCLA,TQUES,TPREP,TCONN,TSPELL,TRELPRPEG I
     6,TSCONJ,TDALE,TENDPT,TTYPE%4<                                     PEG I
      EQUIVALENCE %SUMS%1<,ID<,%SUMS%2<,TITLE<,%SUMS%3<,SENNUM<,%SUMS%4<PEG I
     1,PARNUM<,%SUMS%5<,SUBVER<,%SUMS%6<,SUMLET<,%SUMS%7<,SSQLET<,%SUMS%PEG I
     28<,NUMWDS<,%SUMS%9<,NWDSQ <,              %SUMS%10<,NPAREN<,%SPEG I
     3UMS%11<,NAPOS <,%SUMS%12<,NCOMMA<,%SUMS%13<,NPER  <,%SUMS%14<,NPERPEG I
```

```
     4CT<,%SUMS%15<,NUNDER<,%SUMS%16<,NDASH<,%SUMS%17<,NCOLON<,%SUMS%18 PEG I
     5<,NSEMIC<,%SUMS%19<,NQUOTE<,%SUMS%20<,NEXCLA<,%SUMS%21<,NQUES <,%SPEG I
     6UMS%22<,NPREP<,%SUMS%23<,NCONN<,%SUMS%24<,NSPELL<,%SUMS%25<,NREL  PEG I
     7PR<,%SUMS%26<,NSCONJ<,%SUMS%27<,NDALE<,%SUMS%28<,ENDPCT<,%SUMS%29 PEG I
     8<,SENTYP<                                                         PEG I
      EQUIVALENCE%TOT%1<,TID<,%TOT%2<,TTITLE<,%TOT%3<,NUMSEN<,%TOT%4<,N PEG I
     1UMPAR<,%TOT%5<,SVOPEN<,%TOT%6<,TOTLET<,%TOT%7<,TSQLET<,%TOT%8<,TOTPEG I
     2WDS<,%TOT%9<,TWDSQ<,%TOT%10<,TPAREN<,%TOT%11<,TAP                 PEG I
     3OS<,%TOT%12<,TCOMMA<,%TOT%13<,TPER<,%TOT%14<,TPERCT<,%TOT%15<,TUNDPEG I
     4ER<,%TOT%16<,TDASH<,%TOT%17<,TCOLON<,%TOT%18<,TSEMIC<,%TOT%19<,TQUPEG I
     5OTE<,%TOT%20<,TEXCLA<,%TOT%21<,TQUES<,%TOT%22<,TPREP<,%TOT%23<,TCOPEG I
     6NN<,%TOT%24<,TSPELL<,%TOT%25<,TRELPR<,%TOT%26<,TSCONJ<,%TOT%27<,TDPEG I
     7ALE<,%TOT %28<,TENDPT<,%TOT%29<,TTYPE<                            PEG I
      EQUIVALENCE %SUMS%33<,NHYPH<,%TOT%33<,THYPH<
      EQUIVALENCE  %SUMS%34<,NSLASH<,%TOT%34<,TSLASH<
      INTEGER  NSLASH,TSLASH
      INTEGER  NHYPH,THYPH
      EQUIVALENCE  %NUNDER,NITAL<
      INTEGER  LENGTH %100<
      REAL RDTBL%10540<                                                 PEG I
      EQUIVALENCE %RDTBL%1<,SUBCON<,%RDTBL%41<,PREP<,%RDTBL%141<,RELPRO<PEG I
     1,%RDTBL%161<,CONNEC<,%RDTBL%221<,DALE<,%RDTBL%6221<,SPELLX<,%RDTBLPEG I
     2%10221<,DECLAB<,%RDTBL%10241<,SVOPN<                              PEG I
      REAL HLFTXT%200<
      EQUIVALENCE  %HLFTXT,TEXT<
      COMMON/LOG/SENTND,ESSEND
      LOGICAL SENTND,ESSEND
      EQUIVALENCE%PUNCT%17<,QUOTE<,%PUNCT%18<,PERCT<
      REAL QUOTE,PERCT
      COMMON/PSUM/  RELN,NEXT
      INTEGER  RELN%20<,NEXT
C     RELN %I< CONTAINS THESUMS SUBSCRIPT CORRESPONDING TO PUNCT%I<
      COMMON/LIST2/ SWORD
      DOUBLE PRECISION SWORD%10<
      REAL  SWRD%20<
      EQUIVALENCE  %SWORD,SWRD<
C
C    THIS PART CHECKS A WORD AGAINST A LIST OF COMMONLY MISSPELLED WORDS
C
C     TYPING THE VARIABLE WORD AS DOUBLE PRECISION.
      DOUBLE PRECISION WORD
C     TYPING THE FUNCTION INTABL AS LOGICAL.
      LOGICAL  INTABL
C     A LOGICAL TEST TO DETERMINE WHETHER WORD BEING ANALYZED IS IN
C     THE TABLE OF MISSPELLED WORDS.  IF SO, GO TO 2 FOR INCREMENT-
C     ING.  IF NOT, CONTINUE.
      IF%INTABL%WORD,SPELLX,1430<< GO TO 2
C     RETURN TO MAIN PROGRAM.
      RETURN
C     INCREASE THE NUMBER OF SPELLING WORDS SPELLED INCORRECTLY BY
C     ONE.
2     NSPELL#NSPELL&1
C     RETURN TO MAIN PROGRAM.
      RETURN
      END
$IBFTC TSEN
```

```
      SUBROUTINE  TYPSEN
      COMMON/IN/BROKUP,TEXT,LENGTH                                      PEG I
      REAL BROKUP%80<                                                   PEG I
      DOUBLE PRECISION  TEXT%100<
      COMMON/LISTS/SUBCON,PREP,RELPRO,CONNEC,DALE,SPELLX,DECLAB,SVOPN   PEG I
      DOUBLE PRECISION   SUBCON%20<,PREP%50<,RELPRO%10<,CONNEC%30<,DALE PEG I
     1%3000<,SPELLX%2000<,DECLAB%10<,SVOPN%150<                         PEG I
      COMMON/CHAR/PUNCT,ALPHA                                           PEG I
      REAL PUNCT%20<,BLANK,STAR,DECPT,COMMA,HYPHEN,APOSTR,OPAREN,CPAREN,PEG I
     1SLASH,PERIOD,COLON,SEMIC,EXCLAM,QUEST,DASH,ITALIC,ALPHA%26<,A,B,C,PEG I
     2D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z                     PEG I
      EQUIVALENCE %PUNCT%1<,BLANK<,%PUNCT%2<,STAR<,%PUNCT%3<,DECPT<,%PUNPEG I
     1CT%4<,COMMA<,%PUNCT%5<,HYPHEN<,%PUNCT%6<,APOSTR<,%PUNCT%7<,OPAREN<PEG I
     2,%PUNCT%8<,CPAREN<,%PUNCT%9<,SLASH<,%PUNCT%10<,PERIOD<,%PUNCT%11<,PEG I
     3COLON<,%PUNCT%12<,SEMIC<,%PUNCT%13<,EXCLAM<,%PUNCT%14<,QUEST<,%PUNPEG I
     4CT%15<,DASH<,%PUNCT%16<,ITALIC<,%ALPHA%1<,A<,%ALPHA%2<,B<,%ALPHA%3PEG I
     5<,C<,%ALPHA%4<,D<,%ALPHA%5<,E<,%ALPHA%6<,F<,%ALPHA%7<,G<,%ALPHA%8<PEG I
     6,H<,%ALPHA%9<,I<,%ALPHA%10<,J<,%ALPHA%11<,K<,%ALPHA%12<,L<,%ALPHA%PEG I
     713<,M<,%ALPHA%14<,N<,%ALPHA%15<,O<,%ALPHA%16<,P<,%ALPHA%17<,Q<,%ALPEG I
     8PHA%18<,R<,%ALPHA%19<,S<,%ALPHA%20<,T<,%ALPHA%21<,U<,%ALPHA%22<,V<PEG I
     9,%ALPHA%23<,W<,%ALPHA%24<,X<,%ALPHA%25<,Y<,%ALPHA%26<,Z<          PEG I
      COMMON/OUT/SUMS,TOT                                               PEG I
      INTEGER SUMS%100<,ID,TITLE,SENNUM,PARNUM,SUBVER,SUMLET,SSQLET,NUMWPEG I
     1DS,NWDSQ,NPAREN,NAPOS,NCOMMA,NPER,NPERCT,NUNDER,NDASH,NCOLO       PEG I
     2N,NSEMIC,NQUOTE,NEXCLA,NQUES,NPREP,NCONN,NSPELL,NRELPR,NSCONJ,NDALPEG I
     3E,ENDPCT,SENTYP%4<,TOT%100<,TID,TTITLE,NUMSEN,NUMPAR,SVOPEN,TOTLETPEG I
     4,TSQLET,TOTWDS,TWDSQ,TOTFND,TPAREN,TAPOS,TCOMMA,TPER,TPERCT,TUNDERPEG I
     5,TDASH,TCOLON,TSEMIC,TQUOTE,TEXCLA,TQUES,TPREP,TCONN,TSPELL,TRELPRPEG I
     6,TSCONJ,TDALE,TENDPT,TTYPE%4<                                     PEG I
      EQUIVALENCE %SUMS%1<,ID<,%SUMS%2<,TITLE<,%SUMS%3<,SENNUM<,%SUMS%4<PEG I
     1,PARNUM<,%SUMS%5<,SUBVER<,%SUMS%6<,SUMLET<,%SUMS%7<,SSQLET<,%SUMS%PEG I
     28<,NUMWDS<,%SUMS%9<,NWDSQ <,                %SUMS%10<,NPAREN<,%SPEG I
     3UMS%11<,NAPOS <,%SUMS%12<,NCOMMA<,%SUMS%13<,NPER  <,%SUMS%14<,NPERPEG I
     4CT<,%SUMS%15<,NUNDER<,%SUMS%16<,NDASH<,%SUMS%17<,NCOLON<,%SUMS%18 PEG I
     5<,NSEMIC<,%SUMS%19<,NQUOTE<,%SUMS%20<,NEXCLA<,%SUMS%21<,NQUES <,%SPEG I
     6UMS%22<,NPREP<,%SUMS%23<,NCONN<,%SUMS%24<,NSPELL<,%SUMS%25<,NREL  PEG I
     7PR<,%SUMS%26<,NSCONJ<,%SUMS%27<,NDALE<,%SUMS%28<,ENDPCT<,%SUMS%29 PEG I
     8<,SENTYP<                                                         PEG I
      EQUIVALENCE%TOT%1<,TID<,%TOT%2<,TTITLE<,%TOT%3<,NUMSEN<,%TOT%4<,N PEG I
     1UMPAR<,%TOT%5<,SVOPEN<,%TOT%6<,TOTLET<,%TOT%7<,TSQLET<,%TOT%8<,TOTPEG I
     2WDS<,%TOT%9<,TWDSQ<,%TOT%10<,TPAREN<,%TOT%11<,TAP                 PEG I
     3OS<,%TOT%12<,TCOMMA<,%TOT%13<,TPER<,%TOT%14<,TPERCT<,%TOT%15<,TUNDPEG I
     4ER<,%TOT%16<,TDASH<,%TOT%17<,TCOLON<,%TOT%18<,TSEMIC<,%TOT%19<,TQUPEG I
     5OTE<,%TOT%20<,TEXCLA<,%TOT%21<,TQUES<,%TOT%22<,TPREP<,%TOT%23<,TCOPEG I
     6NN<,%TOT%24<,TSPELL<,%TOT%25<,TRELPR<,%TOT%26<,TSCONJ<,%TOT%27<,TDPEG I
     7ALE<,%TOT %28<,TENDPT<,%TOT%29<,TTYPE<                            PEG I
      EQUIVALENCE %SUMS%33<,NHYPH<,%TOT%33<,THYPH<
      EQUIVALENCE  %SUMS%34<,NSLASH<,%TOT%34<,TSLASH<
      INTEGER  NSLASH,TSLASH
      INTEGER  NHYPH,THYPH
      EQUIVALENCE  %NUNDER,NITAL<
      INTEGER  LENGTH %100<
      REAL RDTBL%10540<                                                 PEG I
      EQUIVALENCE %RDTBL%1<,SUBCON<,%RDTBL%41<,PREP<,%RDTBL%141<,RELPRO<PEG I
     1,%RDTBL%161<,CONNEC<,%RDTBL%221<,DALE<,%RDTBL%6221<,SPELLX<,%RDTBLPEG I
     2%10221<,DECLAB<,%RDTBL%10241<,SVOPN<                              PEG I
```

```
      REAL HLFTXT%200<
      EQUIVALENCE  %HLFTXT,TEXT<
      COMMON/LOG/SENTND,ESSEND
      LOGICAL SENTND,ESSEND
      EQUIVALENCE%PUNCT%17<,QUOTE<,%PUNCT%18<,PERCT<
      REAL QUOTE,PERCT
      COMMON/PSUM/  RELN,NEXT
      INTEGER  RELN%20<,NEXT
C     RELN %I< CONTAINS THESUMS SUBSCRIPT CORRESPONDING TO PUNCT%I<
      COMMON/LIST2/ SWORD
      DOUBLE PRECISION SWORD%10<
      REAL  SWRD%20<
      EQUIVALENCE  %SWORD,SWRD<
      COMMON/ENDS/ ALSPER,ALSEXC,ALSQUS
      REAL  ALSPER%2<,ALSEXC%2<,ALSQUS%2<
      LOGICAL  INTABL
C
C  THIS PART TYPES EACH SENTENCE ACCORDING TO IST END PUNCTUATION
C   AND FIRST WORD TYPE
C
C     REPLACE LAST BY TWO TIMES NEXT MINUS ONE.
      LAST#2*NEXT-1
C     IF SENTENCE ENDS WITH A PERIOD GO TO 100 TO DETERMINE TYPE OF
C     DECLARATIVE SENTENCE.
      IF%HLFTXT%LAST<.EQ.PERIOD.OR.HLFTXT%LAST<.EQ.ALSPER<  GO TO 100
C     IF SENTENCE ENDS WITH AN EXCLAMATION MARK INCREASE SENTENCE
C     TYPE %3< BY ONE.
      IF%HLFTXT%LAST<.EQ.EXCLAM.OR.HLFTXT%LAST<.EQ.ALSEXC<SENTYP%3<#1
C     IF SENTENCE ENDS WITH A QUESTION MARK INCREASE SENTENCE TYPE %4<
C     BY ONE.
      IF%HLFTXT%LAST<.EQ.QUEST .OR.HLFTXT%LAST<.EQ.ALSQUS<SENTYP%4<#1
C     RETURN TO MAIN PROGRAM.
      RETURN
C     TEST TO DETERMINE IF DECLARATIVE SENTENCE IS A PARTICULAR TYPE.
100   IF%INTABL%HLFTXT,DECLAB,20<<  GO TO 102
C     IF 100 FALSE THEN INCREASE DECLARATIVE TYPE A SENTENCE BY ONE.
      SENTYP%1<#1
C     RETURN TO MAIN PROGRAM.
      RETURN
C     IF 100 TRUE INCREASE DECLARATIVE TYPE B SENTENCE BY ONE.
102   SENTYP%2<#1
C     RETURN TO MAIN PROGRAM.
      RETURN
      END
$IBFTC CHK
      SUBROUTINE  CHKLST  %WORD,XX<
      COMMON/IN/BROKUP,TEXT,LENGTH                                     PEG I
      REAL BROKUP%80<                                                  PEG I
      DOUBLE PRECISION  TEXT%100<
      COMMON/LISTS/SUBCON,PREP,RELPRO,CONNEC,DALE,SPELLX,DECLAB,SVOPN  PEG I
      DOUBLE PRECISION   SUBCON%20<,PREP%50<,RELPRO%10<,CONNEC%30<,DALE PEG I
     1%3000<,SPELLX%2000<,DECLAB%10<,SVOPN%150<                        PEG I
      COMMON/CHAR/PUNCT,ALPHA                                          PEG I
      REAL PUNCT%20<,BLANK,STAR,DECPT,COMMA,HYPHEN,APOSTR,OPAREN,CPAREN,PEG I
     1SLASH,PERIOD,COLON,SEMIC,EXCLAM,QUEST,DASH,ITALIC,ALPHA%26<,A,B,C,PEG I
     2D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z                    PEG I
```

```
      EQUIVALENCE %PUNCT%1<,BLANK<,%PUNCT%2<,STAR<,%PUNCT%3<,DECPT<,%PUNPEG I
     1CT%4<,COMMA<,%PUNCT%5<,HYPHEN<,%PUNCT%6<,APOSTR<,%PUNCT%7<,OPAREN<PEG I
     2,%PUNCT%8<,CPAREN<,%PUNCT%9<,SLASH<,%PUNCT%10<,PERIOD<,%PUNCT%11<,PEG I
     3COLON<,%PUNCT%12<,SEMIC<,%PUNCT%13<,EXCLAM<,%PUNCT%14<,QUEST<,%PUNPEG I
     4CT%15<,DASH<,%PUNCT%16<,ITALIC<,%ALPHA%1<,A<,%ALPHA%2<,B<,%ALPHA%3PEG I
     5<,C<,%ALPHA%4<,D<,%ALPHA%5<,E<,%ALPHA%6<,F<,%ALPHA%7<,G<,%ALPHA%8<PEG I
     6,H<,%ALPHA%9<,I<,%ALPHA%10<,J<,%ALPHA%11<,K<,%ALPHA%12<,L<,%ALPHA%PEG I
     713<,M<,%ALPHA%14<,N<,%ALPHA%15<,O<,%ALPHA%16<,P<,%ALPHA%17<,Q<,%ALPEG I
     8PHA%18<,R<,%ALPHA%19<,S<,%ALPHA%20<,T<,%ALPHA%21<,U<,%ALPHA%22<,V<PEG I
     9,%ALPHA%23<,W<,%ALPHA%24<,X<,%ALPHA%25<,Y<,%ALPHA%26<,Z<             PEG I
      COMMON/OUT/SUMS,TOT                                                  PEG I
      INTEGER SUMS%100<,ID,TITLE,SENNUM,PARNUM,SUBVER,SUMLET,SSQLET,NUMWPEG I
     1DS,NWDSQ,NPAREN,NAPOS,NCOMMA,NPER,NPERCT,NUNDER,NDASH,NCOLO          PEG I
     2N,NSEMIC,NQUOTE,NEXCLA,NQUES,NPREP,NCONN,NSPELL,NRELPR,NSCONJ,NDALPEG I
     3E,ENDPCT,SENTYP%4<,TOT%100<,TID,TTITLE,NUMSEN,NUMPAR,SVOPEN,TOTLETPEG I
     4,TSQLET,TOTWDS,TWDSQ,TOTFND,TPAREN,TAPOS,TCOMMA,TPER,TPERCT,TUNDERPEG I
     5,TDASH,TCOLON,TSEMIC,TQUOTE,TEXCLA,TQUES,TPREP,TCONN,TSPELL,TRELPRPEG I
     6,TSCONJ,TDALE,TENDPT,TTYPE%4<                                        PEG I
      EQUIVALENCE %SUMS%1<,ID<,%SUMS%2<,TITLE<,%SUMS%3<,SENNUM<,%SUMS%4<PEG I
     1,PARNUM<,%SUMS%5<,SUBVER<,%SUMS%6<,SUMLET<,%SUMS%7<,SSQLET<,%SUMS%PEG I
     28<,NUMWDS<,%SUMS%9<,NWDSQ <,                   %SUMS%10<,NPAREN<,%SPEG I
     3UMS%11<,NAPOS <,%SUMS%12<,NCOMMA<,%SUMS%13<,NPER  <,%SUMS%14<,NPERPEG I
     4CT<,%SUMS%15<,NUNDER<,%SUMS%16<,NDASH<,%SUMS%17<,NCOLON<,%SUMS%18 PEG I
     5<,NSEMIC<,%SUMS%19<,NQUOTE<,%SUMS%20<,NEXCLA<,%SUMS%21<,NQUES <,%SPEG I
     6UMS%22<,NPREP<,%SUMS%23<,NCONN<,%SUMS%24<,NSPELL<,%SUMS%25<,NREL  PEG I
     7PR<,%SUMS%26<,NSCONJ<,%SUMS%27<,NDALE<,%SUMS%28<,ENDPCT<,%SUMS%29 PEG I
     8<,SENTYP<                                                            PEG I
      EQUIVALENCE%TOT%1<,TID<,%TOT%2<,TTITLE<,%TOT%3<,NUMSEN<,%TOT%4<,N PEG I
     1UMPAR<,%TOT%5<,SVOPEN<,%TOT%6<,TOTLET<,%TOT%7<,TSQLET<,%TOT%8<,TOTPEG I
     2WDS<,%TOT%9<,TWDSQ<,%TOT%10<,TPAREN<,%TOT%11<,TAP                    PEG I
     3OS<,%TOT%12<,TCOMMA<,%TOT%13<,TPER<,%TOT%14<,TPERCT<,%TOT%15<,TUNDPEG I
     4ER<,%TOT%16<,TDASH<,%TOT%17<,TCOLON<,%TOT%18<,TSEMIC<,%TOT%19<,TQUPEG I
     5OTE<,%TOT%20<,TEXCLA<,%TOT%21<,TQUES<,%TOT%22<,TPREP<,%TOT%23<,TCOPEG I
     6NN<,%TOT%24<,TSPELL<,%TOT%25<,TRELPR<,%TOT%26<,TSCONJ<,%TOT%27<,TDPEG I
     7ALE<,%TOT %28<,TENDPT<,%TOT%29<,TTYPE<                               PEG I
      EQUIVALENCE %SUMS%33<,NHYPH<,%TOT%33<,THYPH<
      EQUIVALENCE  %SUMS%34<,NSLASH<,%TOT%34<,TSLASH<
      INTEGER  NSLASH,TSLASH
      INTEGER  NHYPH,THYPH
      EQUIVALENCE  %NUNDER,NITAL<
      INTEGER  LENGTH %100<
                                                                           PEG I
      REAL RDTBL%10540<
      EQUIVALENCE %RDTBL%1<,SUBCON<,%RDTBL%41<,PREP<,%RDTBL%141<,RELPRO<PEG I
     1,%RDTBL%161<,CONNEC<,%RDTBL%221<,DALE<,%RDTBL%6221<,SPELLX<,%RDTBLPEG I
     2%10221<,DECLAB<,%RDTBL%10241<,SVOPN<                                 PEG I
      REAL HLFTXT%200<
      EQUIVALENCE  %HLFTXT,TEXT<
      COMMON/LOG/SENTND,ESSEND
      LOGICAL SENTND,ESSEND
      EQUIVALENCE%PUNCT%17<,QUOTE<,%PUNCT%18<,PERCT<
      REAL QUOTE,PERCT
      COMMON/PSUM/  RELN,NEXT
      INTEGER  RELN%20<,NEXT
C     RELN %I< CONTAINS THESUMS SUBSCRIPT CORRESPONDING TO PUNCT%I<
      COMMON/LIST2/ SWORD
      DOUBLE PRECISION SWORD%10<
```

APPENDIX A (Continued)

```
      REAL  SWRD%20<
      EQUIVALENCE  %SWORD,SWRD<
C
C  THIS PART CLASSIFIES WORDS OF THE SENTENCE AS PREPOSITION,
C     RELATIVE PRONOUNS, SUBORDINATING CONJUNCTIONS CONNECTIVES,
C  AND/OR ONE OF THE 3000  COMMON WORDS ON THE DALE LIST
C
C     TYPING WORD AS DOUBLE PRECISION, I.E. IT CAN CONTAIN FROM 1-12
C     CHARACTERS.
      DOUBLE PRECISION WORD
C     TYPING XX AND YY AS LOGICAL VARIABLES.
      LOGICAL XX,YY
C     TYPING THE FUNCTION INTABL AS LOGICAL.
      LOGICAL  INTABL
C     REPLACING THE VALUE OF YY BY THE VALUE OF XX.
      YY#XX
C     REPLACING XX BY THE LOGICAL CONSTANT TRUE.
      XX#.TRUE.
C     TEST TO DETERMINE IF THE WORD IS A PREPOSITION.
      IF%INTABL%WORD,PREP,100<<  GO TO 100
C     YY WILL BE TRUE FOR FIRST AND SECOND WORD OF THE SENTENCE.
C     AVOIDS CHECKING THESE WORDS AS BEING RELATIVE PRONOUNS.
      IF%YY<  GO TO 6
C     TEST TO DETERMINE IF THE WORD IS A RELATIVE PRONOUN.
      IF%INTABL%WORD,RELPRO,20<< GO TO 200
C     TEST TO DETERMINE IF THE WORD IS A SUBORDINATE CONJUNCTION.
6     IF%INTABL%WORD,SUBCON,40<<  GO TO 300
C     TEST TO DETERMINE IF THE WORD IS IN THE LIST OF DALE WORDS.
      IF%INTABL%WORD,DALE,6000<<  GO TO 400
C     TEST TO DETERMINE IF THE WORD IS A CONNECTIVE.
      IF%INTABL%WORD,CONNEC,60<<  GO TO 500
C     ASSIGN THE LOGICAL CONSTANT FALSE TO XX.
      XX#.FALSE.
C     RETURN TO CALLING PROGRAM.
      RETURN
C     INCREMENT NUMBER OF PREPOSITIONS BY ONE.
100   NPREP#NPREP&1
C     ASSIGNED GO TO FOR ANOTHER INCREMENT.
      GO TO 400
C     INCREMENT NUMBER OF RELATIVE PRONOUNS BY ONE.
200   NRELPR#NRELPR&1
C     ASSIGNED GO TO FOR ANOTHER INCREMENT.
      GO TO 400
C     INCREMENT NUMBER OF SUBORDINATE CONJUNCTIONS BY ONE.
300   NSCONJ#NSCONJ&1
C     INCREMENT THE NUMBER OF DALE WORDS BY ONE.
400   NDALE#NDALE&1
C     TEST TO DETERMINE IF THE WORD IS A CONNECTIVE.
      IF%INTABL%WORD,CONNEC,60<< GO TO 500
C     RETURN TO CALLING PROGRAM.
      RETURN
C     INCREMENT NUMBER OF CONNECTIVES BY ONE.
500   NCONN#NCONN&1
C     RETURN TO CALLING PROGRAM.
      RETURN
      END
```

```
$IBFTC SYB
      SUBROUTINE  SEEOPN
      COMMON/IN/BROKUP,TEXT,LENGTH                                      PEG I
      REAL BROKUP%80<                                                   PEG I
      DOUBLE PRECISION  TEXT%100<
      COMMON/LISTS/SUBCON,PREP,RELPRO,CONNEC,DALE,SPELLX,DECLAB,SVOPN   PEG I
      DOUBLE PRECISION   SUBCON%20<,PREP%50<,RELPRO%10<,CONNEC%30<,DALE PEG I
     1%3000<,SPELLX%2000<,DECLAB%10<,SVOPN%150<                         PEG I
      COMMON/CHAR/PUNCT,ALPHA                                           PEG I
      REAL PUNCT%20<,BLANK,STAR,DECPT,COMMA,HYPHEN,APOSTR,OPAREN,CPAREN,PEG I
     1SLASH,PERIOD,COLON,SEMIC,EXCLAM,QUEST,DASH,ITALIC,ALPHA%26<,A,B,C,PEG I
     2D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z                     PEG I
      EQUIVALENCE %PUNCT%1<,BLANK<,%PUNCT%2<,STAR<,%PUNCT%3<,DECPT<,%PUNPEG I
     1CT%4<,COMMA<,%PUNCT%5<,HYPHEN<,%PUNCT%6<,APOSTR<,%PUNCT%7<,OPAREN<PEG I
     2,%PUNCT%8<,CPAREN<,%PUNCT%9<,SLASH<,%PUNCT%10<,PERIOD<,%PUNCT%11<,PEG I
     3COLON<,%PUNCT%12<,SEMIC<,%PUNCT%13<,EXCLAM<,%PUNCT%14<,QUEST<,%PUNPEG I
     4CT%15<,DASH<,%PUNCT%16<,ITALIC<,%ALPHA%1<,A<,%ALPHA%2<,B<,%ALPHA%3PEG I
     5<,C<,%ALPHA%4<,D<,%ALPHA%5<,E<,%ALPHA%6<,F<,%ALPHA%7<,G<,%ALPHA%8<PEG I
     6,H<,%ALPHA%9<,I<,%ALPHA%10<,J<,%ALPHA%11<,K<,%ALPHA%12<,L<,%ALPHA%PEG I
     713<,M<,%ALPHA%14<,N<,%ALPHA%15<,O<,%ALPHA%16<,P<,%ALPHA%17<,Q<,%ALPEG I
     8PHA%18<,R<,%ALPHA%19<,S<,%ALPHA%20<,T<,%ALPHA%21<,U<,%ALPHA%22<,V<PEG I
     9,%ALPHA%23<,W<,%ALPHA%24<,X<,%ALPHA%25<,Y<,%ALPHA%26<,Z<          PEG I
      COMMON/OUT/SUMS,TOT                                               PEG I
      INTEGER SUMS%100<,ID,TITLE,SENNUM,PARNUM,SUBVER,SUMLET,SSQLET,NUMWPEG I
     1DS,NWDSQ,NPAREN,NAPOS,NCOMMA,NPER,NPERCT,NUNDER,NDASH,NCOLO       PEG I
     2N,NSEMIC,NQUOTE,NEXCLA,NQUES,NPREP,NCONN,NSPELL,NRELPR,NSCONJ,NDALPEG I
     3E,ENDPCT,SENTYP%4<,TOT%100<,TID,TTITLE,NUMSEN,NUMPAR,SVOPEN,TOTLETPEG I
     4,TSQLET,TOTWDS,TWDSQ,TOTFND,TPAREN,TAPOS,TCOMMA,TPER,TPERCT,TUNDERPEG I
     5,TDASH,TCOLON,TSEMIC,TQUOTE,TEXCLA,TQUES,TPREP,TCONN,TSPELL,TRELPRPEG I
     6,TSCONJ,TDALE,TENDPT,TTYPE%4<                                     PEG I
      EQUIVALENCE %SUMS%1<,ID<,%SUMS%2<,TITLE<,%SUMS%3<,SENNUM<,%SUMS%4<PEG I
     1,PARNUM<,%SUMS%5<,SUBVER<,%SUMS%6<,SUMLET<,%SUMS%7<,SSQLET<,%SUMS%PEG I
     28<,NUMWDS<,%SUMS%9<,NWDSQ <,               %SUMS%10<,NPAREN<,%SPEG I
     3UMS%11<,NAPOS <,%SUMS%12<,NCOMMA<,%SUMS%13<,NPER  <,%SUMS%14<,NPERPEG I
     4CT<,%SUMS%15<,NUNDER<,%SUMS%16<,NDASH<,%SUMS%17<,NCOLON<,%SUMS%18 PEG I
     5<,NSEMIC<,%SUMS%19<,NQUOTE<,%SUMS%20<,NEXCLA<,%SUMS%21<,NQUES <,%SPEG I
     6UMS%22<,NPREP<,%SUMS%23<,NCONN<,%SUMS%24<,NSPELL<,%SUMS%25<,NREL  PEG I
     7PR<,%SUMS%26<,NSCONJ<,%SUMS%27<,NDALE<,%SUMS%28<,ENDPCT<,%SUMS%29 PEG I
     8<,SENTYP<                                                         PEG I
      EQUIVALENCE%TOT%1<,TID<,%TOT%2<,TTITLE<,%TOT%3<,NUMSEN<,%TOT%4<,N PEG I
     1UMPAR<,%TOT%5<,SVOPEN<,%TOT%6<,TOTLET<,%TOT%7<,TSQLET<,%TOT%8<,TOTPEG I
     2WDS<,%TOT%9<,TWDSQ<,%TOT%10<,TPAREN<,%TOT%11<,TAP                 PEG I
     3OS<,%TOT%12<,TCOMMA<,%TOT%13<,TPER<,%TOT%14<,TPERCT<,%TOT%15<,TUNDPEG I
     4ER<,%TOT%16<,TDASH<,%TOT%17<,TCOLON<,%TOT%18<,TSEMIC<,%TOT%19<,TQUPEG I
     5OTE<,%TOT%20<,TEXCLA<,%TOT%21<,TQUES<,%TOT%22<,TPREP<,%TOT%23<,TCOPEG I
     6NN<,%TOT%24<,TSPELL<,%TOT%25<,TRELPR<,%TOT%26<,TSCONJ<,%TOT%27<,TDPEG I
     7ALE<,%TOT %28<,TENDPT<,%TOT%29<,TTYPE<                            PEG I
      EQUIVALENCE %SUMS%33<,NHYPH<,%TOT%33<,THYPH<
      EQUIVALENCE  %SUMS%34<,NSLASH<,%TOT%34<,TSLASH<
      INTEGER  NSLASH,TSLASH
      INTEGER  NHYPH,THYPH
      EQUIVALENCE  %NUNDER,NITAL<
      INTEGER  LENGTH %100<
      REAL RDTBL%10540<                                                 PEG I
      EQUIVALENCE %RDTBL%1<,SUBCON<,%RDTBL%41<,PREP<,%RDTBL%141<,RELPRO<PEG I
     1,%RDTBL%161<,CONNEC<,%RDTBL%221<,DALE<,%RDTBL%6221<,SPELLX<,%RDTBLPEG I
```

```
      2%10221<,DECLAB<,%RDTBL%10241<,SVOPN<                              PEG I
        REAL HLFTXT%200<
        EQUIVALENCE  %HLFTXT,TEXT<
        COMMON/LOG/SENTND,ESSEND
        LOGICAL SENTND,ESSEND
        EQUIVALENCE%PUNCT%17<,QUOTE<,%PUNCT%18<,PERCT<
        REAL QUOTE,PERCT
        COMMON/PSUM/  RELN,NEXT
        INTEGER  RELN%20<,NEXT
C      RELN %I< CONTAINS THESUMS SUBSCRIPT CORRESPONDING TO PUNCT%I<
        COMMON/LIST2/ SWORD
        DOUBLE PRECISION SWORD%10<
        REAL  SWRD%20<
        EQUIVALENCE  %SWORD,SWRD<
        REAL  ENDING,WRDEND%2<
C
C  THIS PART CLASSIFIES THE SENTENCE ACCORDING TO WHTHER THE OPENING
C  IS A SUBJECT VERB TYPE OPENING OR NOT
C    CONSIST OF  CHECKING FOR CERTAIN  WORDS , ELIMINATING CERTAIN WORDS
C   AND  TESTING THE WORD ENDING OF THE FIRST WORD OF THE SENTENCE
C  FOR S
C
C       TYPING THE ARRAY  LETTER  AS REAL.
        REAL  LETTER%12<
C       TYPING THE FUNCTION  INTABL  AS LOGICAL.
        LOGICAL  INTABL
C       INITIALIZE SUBJECT-VERB VARIABLE.
        SUBVER#0
C       TEST TO DETERMINE IF FIRST WORD ENDS IN A  S, IF SO, RETURN TO
C       CALLING PROGRAM.
        IF%INTABL%TEXT,SWORD,20<<  RETURN
C       REPLACE  ICT  BY  LENGTH%1<.
        ICT#LENGTH%1<
C       REPLACE  LL  BY ICT MINUS FIVE.
        LL#ICT-5
C       UNPACK WORD.
        CALL UNPACK%TEXT%1<,LETTER,12<
C       PACKS  WRDEND  WITH WORD ENDING DESIGNATED BY LETTER.
        CALL PACK %LETTER%LL<,WRDEND,6<
C       THIS CALL STATEMENT SETS  ENDING  EQUAL TO  ATION.
        CALL DATA %ENDING,6HATION <
C       TEST TO DETERMINE IF  WRDEND  IS SAME AS  ATION.
        IF % ENDING .EQ. WRDEND < GO TO 3
C       THIS CALL STATEMENT SETS  ENDING  EQUAL TO  OLOGY.
        CALL DATA %ENDING,6HOLOGY <
C       TEST TO DETERMINE IF  WRDEND  IS SAME AS  OLOGY.
        IF % ENDING .EQ. WRDEND < GO TO 3
C       THIS CALL STATEMENT SETS  ENDING  EQUAL TO  SHIP.
        CALL DATA %ENDING,6HSHIP  <
C       TEST TO DETERMINE IF  WRDEND  IS SAME AS  SHIP.
        IF % ENDING .EQ. WRDEND < GO TO 3
C       THIS CALL STATEMENT SETS  ENDING  EQUAL TO  MENT.
        CALL DATA %ENDING,6HMENT  <
C       TEST TO DETERMINE IF  WRDEND  IS SAME AS  MENT.
        IF % ENDING .EQ. WRDEND < GO TO 3
C       TEST TO DETERMINE IF WORD BEING ANALYZED IS IN S-V LIST.
```

```
      IF%INTABL%TEXT%1<,SVOPN,300<<  GO TO 3
C     UNPACK LETTER.
      CALL UNPACK  %TEXT%1<,LETTER,12<
C     IF LAST LETTER NOT EQUAL TO  S  RETURN TO CALLING PROGRAM.
      IF%LETTER%ICT<.NE.S<  RETURN
C     IF WORD IS IN S-V LIST, HAS ONE OF THE ABOVE ENDINGS, OR ENDS
C     WITH S, THEN CHANGE VARIABLE SVBVER FROM 0 TO 1 INDICATING
C     A  S-V  OPENING.
3     SUBVER#1
C     RETURN TO CALLING PROGRAM.
      RETURN
      END
```

```
$IBFTC TLU
      LOGICAL FUNCTION INTABL%A,B,M<
      COMPLEX A,B%8192<
      INTEGER N
C     LOGICAL COMPARISON SUBROUTINES.
      LOGICAL EQA,GTA,LTA
C
C BINARY SEARCH--A IS THE ARGUMENT,B THE TABLE,N THE TABLE LENGTH.
C
C     REPLACE THE VALUE OF N BY THE VALUE IN  M  DIVIDED BY TWO.
      N#M/2
C     REPLACE  INTABL  BY THE LOGICAL CONSTANT FALSE.
      INTABL#.FALSE.
C     REPLACE THE VARIABLE  J  BY THE CONSTANT 4096.
      J#4096
C     REPLACE THE VALUE OF  K  BY THE VALUE OF  J.
      K#J
C     TEST TO DETERMINE IF  J  EQUAL TO ZERO.  IF SO, RETURN TO
C     CALLING PROGRAM.
1     IF%J.EQ.0< RETURN
C     REPLACE THE VALUE OF  J  BY THE VALUE OF  J  DIVIDED BY TWO.
      J#J/2
C     REPLACE  L  BY THE MINIMUM VALUE OF THE TWO ARGUMENTS,
C     K  OR  N.
      L#MINO%K,N<
      IF%LTA%REAL%A<,REAL%B%L<<<< GO TO 3
      IF%GTA%REAL%A<,REAL%B%L<<<< GO TO 2
      IF%LTA%AIMAG%A<,AIMAG%B%L<<<< GO TO 3
      IF%GTA%AIMAG%A<,AIMAG%B%L<<<< GO TO 2
C     REPLACE  INTABL  BY THE LOGICAL CONSTANT TRUE.
      INTABL#.TRUE.
C     RETURN TO CALLING PROGRAM.
      RETURN
C     REPLACE  K  BY THE VALUE IN  K  PLUS THE VALUE IN  J.
2     K#K&J
C     ASSIGNED GO TO FOR A TEST.
      GO TO 1
C     REPLACE  K  BY THE VALUE IN  K  MINUS THE VALUE IN  J.
3     K#K-J
C     ASSIGN GO TO FOR A TEST.
      GO TO 1
      END
$IBFTC SUBUPC
      SUBROUTINE UNPACK%A,B,N<
```

```
      REAL   LSHIFT
      DIMENSION  A%4<,B%100<
      DO  1 I#1,N
      J# %I-1</6&1
      K#MOD%I-1,6<
1     B%I<#LSHIFT%A%J<,K<
      RETURN
      END
$IBMAP LSHFT
       ENTRY    LSHIFT
LSHIFT SAVE     4
       CLA*     4,4
       ADD*     4,4
       STO      TEMP
       ADD      TEMP
       ADD      TEMP
       STA      S
       CAL*     3,4
S      ALS      **
       ANA      #07700000000000
       ORA      #06060606060
       SLW      TEMP
       CLA      TEMP
       RETURN   LSHIFT
TEMP   BSS      1
       END
$IBFTC SUBPAC
       SUBROUTINE  PACK%A,B,N<
C     TYPING RSHIFT AS REAL.
      REAL   RSHIFT
C     DIMENSIONING THE ARRAYS  A  AND  B.
      DIMENSION  A%100<,B%4<
C     N  DETERMINED BY ARGUMENT VALUE IN CALL STATEMENT.
      M#%%N&11</12<*12
      DO 1 I#1,M
      J#%I-1</6&1
C     MODULO FUNCTION WHEREIN INTERESTED IN REMAINDER OF I-1 DIVIDED
C     BY 6.
      K#MOD%I-1,6<
      IF%K.EQ.0<  B%J<#0.
      D#A%I<
      IF%I.GT.N<  D#-.5
1     B%J<#OR%B%J<,RSHIFT%D,K<<
      RETURN
      END
$IBMAP RSHFT
       ENTRY    RSHIFT
RSHIFT SAVE     4
       STQ      HOLD
       CLA*     4,4
       ADD*     4,4
       STO      TEMP
       ADD      TEMP
       ADD      TEMP
       STA      S
       CAL*     3,4
```

```
        ANA     #0770000000000
S       LGR     **
        SLW     TEMP
        CLA     TEMP
        LDQ     HOLD
        RETURN  RSHIFT
TEMP    BSS     1
HOLD    BSS     1
        END
$IBFTC DATSIN
       SUBROUTINE DATA%A,B<
C      TYPING THE ARGUMENTS  A  AND  B  AS REAL.
       REAL A,B
C      SET ONE ARGUMENT EQUAL TO THE OTHER.
       A#B
C      RETURN TO CALLING PROGRAM.
       RETURN
       END
```

```
$IBMAP LCMP
*
*    LOGICAL COMPARISON SUBROUTINES
*
        ENTRY   GTA
GTA     SAVE    4
        CAL*    3,4
        LAS*    4,4
        TRA     RA1
        NOP
        ZAC
        RETURN  GTA
RA1     CLS     #0
        RETURN  GTA
        ENTRY   GEA
GEA     SAVE    4
        CAL*    3,4
        LAS*    4,4
        TRA     RA2
        TRA     RA2
        ZAC
        RETURN  GEA
RA2     CLS     #0
        RETURN  GEA
        ENTRY   LTA
LTA     SAVE    4
        CAL*    4,4
        LAS*    3,4
        TRA     RA3
        NOP
        ZAC
        RETURN  LTA
RA3     CLS     #0
        RETURN  LTA
        ENTRY   LEA
LEA     SAVE    4
        CAL*    4,4
        LAS*    3,4
```

```
          TRA      RA4
          TRA      RA4
          ZAC
          RETURN   LEA
RA4       CLS      #0
          RETURN   LEA
          ENTRY    EQA
EQA       SAVE     4
          CAL*     3,4
          LAS*     4,4
          TRA      *&2
          TRA      RA5
          ZAC
          RETURN   EQA
RA5       CLS      #0
          RETURN   EQA
          ENTRY    NEA
NEA       SAVE     4
          CAL*     3,4
          LAS*     4,4
          TRA      *&2
          TRA      RA6
          CLS      #0
          RETURN   NEA
RA6       ZAC
          RETURN   NEA
          END
$IBMAP    SETFP
          ENTRY    FPTRAP
FPTRAP    SAVE     4
          AXT      -1,4
          SXA      SETFP.&14,4
          RETURN   FPTRAP
          EXTERN   SETFP.
          END
$ENTRY             PEGI

*
.
,
-

%
<
/
.
..
.,
.X
.Q
--
%/<
*
%
```

A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
U
V
W
X
Y
Z

APPENDIX A (Continued)

AFTER	ALTHOUGH	AS	BECAUSE	BEFORE	IF
SINCE	THAN	THOUGH	UNLESS	UNTIL	WHENEVER
WHEN	WHERE	WHEREVER	WHILE	ZZZZZZZZZZZZZZZZZZZZZZZZZ	
ZZZZZZZZZZZZZZZZZZZZZZZZZ		ABOUT	ABOVE	ACROSS	AGAINST
ALONG	AMID	AMONG	AROUND	AT	BEHIND
BELOW	BENEATH	BESIDE	BETWEEN	BEYOND	BUT
BY	CONCERNING	DOWN	DURING	EXCEPT	FOR
FROM	IN	INTO	OFF	OF	ON
OVER	PAST	THROUGHOUT	THROUGH	TOWARD	UNDERNEATH
UNDER	UNTO	UPON	UP	WITHIN	WITHOUT
WITH	ZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZ				
ZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZ				THAT	WHICH
WHOM	WHO	WHOSE	ZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZ		
ZZZZZZZZZZZZZZZZZZZZZZZZZ		ACCORDINGLY	AFTERWARD	CONSEQUENTLY	FIFTH
FINALLY	FIRST	FOURTH	FURTHERMORE	HENCE	HOWEVER
INDEED	INSTEAD	LIKEWISE	MEANWHILE	MOREOVER	NEVERTHELESS
OTHERWISE	SECOND	SIMILARLY	THEREFORE	THIRD	THUS
YET	ZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZ				
ZZZZZZZZZZZZZZZZZZZZZZZZZ		TIS	ABLE	ABOARD	ABOUT
ABOVE	ABSENT	ACCEPT	ACCIDENT	ACCOUNT	ACHE
ACHING	ACORN	ACRE	ACROSS	ACT	ACTS
ADDRESS	ADD	ADMIRE	ADVENTURE	AFAR	AFRAID
AFTERNOON	AFTER	AFTERWARD	AFTERWARDS	AGAIN	AGAINST
AGED	AGE	AGO	AGREE	AHEAD	AH
AID	AIM	AIRFIELD	AIRPLANE	AIRPORT	AIR
AIRSHIP	AIRY	ALARM	ALIKE	ALIVE	ALLEY
ALLIGATOR	ALLOW	ALL	ALMOST	ALONE	ALONG
ALOUD	ALREADY	ALSO	ALWAYS	AMERICAN	AMONG
AMOUNT	AM	AND	ANGEL	ANGER	ANGRY
ANIMAL	ANOTHER	AN	ANSWER	ANT	ANYBODY

APPENDIX A (Continued)

ANYHOW	ANYONE	ANY	ANYTHING	ANYWAY	ANYWHERE
APARTMENT	APART	APE	APIECE	APPEAR	APPLE
APRIL	APRON	AREN T	ARE	ARISE	ARITHMETIC
ARMFUL	ARM	ARMY	AROSE	AROUND	ARRANGE
ARRIVED	ARRIVE	ARROW	ARTIST	ART	A
ASHES	ASH	ASIDE	ASK	ASLEEP	AS
ATE	AT	ATTACK	ATTEND	ATTENTION	AUGUST
AUNT	AUTHOR	AUTOMOBILE	AUTO	AUTUMN	AVENUE
AWAKEN	AWAKE	AWAY	AWFULLY	AWFUL	AWHILE
AX	BAA	BABE	BABIES	BABY	BACKGROUND
BACK	BACKWARD	BACKWARDS	BACON	BADGE	BADLY
BAD	BAG	BAKER	BAKERY	BAKE	BAKING
BALLOON	BALL	BANANA	BANDAGE	BAND	BANG
BANJO	BANKER	BANK	BARBER	BAREFOOT	BARELY
BARE	BARK	BARN	BARREL	BAR	BASEBALL
BASEMENT	BASE	BASKET	BATCH	BATHE	BATHING
BATHROOM	BATH	BATHTUB	BAT	BATTLE	BATTLESHIP
BAY	BEACH	BEAD	BEAM	BEAN	BEARD
BEAR	BEAST	BEATING	BEAT	BEAUTIFUL	BEAUTIFY
BEAUTY	BECAME	BECAUSE	BECOME	BECOMING	BEDBUG
BEDROOM	BED	BEDSPREAD	BEDTIME	BEECH	BEEF
BEEHIVE	BEEN	BEER	BEE	BEESTEAK	BEET
BEFORE	BEGAN	BEGGAR	BEGGED	BEGINNING	BEGIN
BEG	BEGUN	BEHAVE	BEHIND	BEING	BELIEVE
BELL	BELONG	BELOW	BELT	BENCH	BEND
BENEATH	BENT	BERRIES	BERRY	BE	BESIDE
BESIDES	BEST	BET	BETTER	BETWEEN	BIBLE
BIB	BICYCLE	BID	BIGGER	BIG	BILLBOARD
BILL	BIND	BIN	BIRD	BIRTHDAY	BIRTH
BISCUIT	BITE	BITING	BIT	BITTER	BLACKBERRY
BLACKBIRD	BLACKBOARD	BLACKNESS	BLACK	BLACKSMITH	BLAME
BLANKET	BLANK	BLAST	BLAZE	BLEED	BLESSING
BLESS	BLEW	BLINDFOLD	BLIND	BLINDS	BLOCK
BLOOD	BLOOM	BLOSSOM	BLOT	BLOW	BLUEBERRY
BLUEBIRD	BLUEJAY	BLUE	BLUSH	BOARD	BOAST
BOAT	BOB	BOBWHITE	BODIES	BODY	BOILER
BOIL	BOLD	BONE	BONNET	BOOKCASE	BOOKKEEPER
BOOK	BOOM	BOO	BOOT	BORN	BORROW
BOSS	BOTHER	BOTH	BOTTLE	BOTTOM	BOUGHT
BOUNCE	BOW-WOW	BOWL	BOW	BOXCAR	BOXER
BOXES	BOX	BOYHOOD	BOY	BRACELET	BRAIN
BRAKE	BRANCH	BRAN	BRASS	BRAVE	BREAD
BREAKFAST	BREAK	BREAST	BREATHE	BREATH	BREEZE
BRICK	BRIDE	BRIDGE	BRIGHTNESS	BRIGHT	BRING
BROADCAST	BROAD	BROKEN	BROKE	BROOK	BROOM
BROTHER	BROUGHT	BROWN	BRUSH	BUBBLE	BUCKET
BUCKLE	BUD	BUFFALO	BUGGY	BUG	BUILDING
BUILD	BUILT	BULB	BULLET	BULL	BUMBLEBEE
BUMP	BUM	BUNCH	BUNDLE	BUNNY	BUN
BURN	BURST	BURY	BUSHEL	BUSH	BUSINESS
BUS	BUSY	BUTCHER	BUT	BUTTERCUP	BUTTERFLY
BUTTERMILK	BUTTER	BUTTERSCOTCH	BUTTONHOLE	BUTTON	BUTT
BUY	BUZZ	BYE	BY	CABBAGE	CABINET
CABIN	CAB	CACKLE	CAGE	CAKE	CALENDAR
CALF	CALLER	CALLING	CALL	CAMEL	CAME
CAMPFIRE	CAMP	CAN T	CANAL	CANARY	CANDLE

CANDLESTICK	CANDY	CANE	CANNON	CANNOT	CANOE
CAN	CANYON	CAPE	CAPITAL	CAP	CAPTAIN
CARDBOARD	CARD	CAREFUL	CARELESSNESS	CARELESS	CARE
CARLOAD	CARPENTER	CARPET	CARRIAGE	CARROT	CARRY
CAR	CART	CARVE	CASE	CASHIER	CASH
CASTLE	CATBIRD	CATCHER	CATCH	CATERPILLAR	CATFISH
CAT	CATSUP	CATTLE	CAUGHT	CAUSE	CAVE
CEILING	CELLAR	CELL	CENTER	CENT	CEREAL
CERTAINLY	CERTAIN	CHAIN	CHAIR	CHALK	CHAMPION
CRY	CUB	CUFF	CUPBOARD	CUPFUL	CUP
CHANCE	CHANGE	CHAP	CHARGE	CHARM	CHART
CHASE	CHATTER	CHEAP	CHEAT	CHECKERS	CHECK
CHEEK	CHEER	CHEESE	CHERRY	CHEST	CHEW
CHICKEN	CHICK	CHIEF	CHILDHOOD	CHILDREN	CHILD
CHILL	CHILLY	CHIMNEY	CHINA	CHIN	CHIPMUNK
CHIP	CHOCOLATE	CHOICE	CHOOSE	CHOP	CHORUS
CHOSEN	CHOSE	CHRISTEN	CHRISTMAS	CHURCH	CHURN
CIGARETTE	CIRCLE	CIRCUS	CITIZEN	CITY	CLANG
CLAP	CLASSMATE	CLASSROOM	CLASS	CLAW	CLAY
CLEANER	CLEAN	CLEAR	CLERK	CLEVER	CLICK
CLIFF	CLIMB	CLIP	CLOAK	CLOCK	CLOSE
CLOSET	CLOTHES	CLOTHING	CLOTH	CLOUD	CLOUDY
CLOVER	CLOWN	CLUB	CLUCK	CLUMP	COACH
COAL	COAST	COAT	COBBLER	COB	COCOA
COCONUT	COCOON	CODFISH	COD	COFFEEPOT	COFFEE
COIN	COLD	COLLAR	COLLEGE	COLORED	COLOR
COLT	COLUMN	COMB	COME	COMFORT	COMIC
COMING	COMPANY	COMPARE	CONDUCTOR	CONE	CONNECT
COOKED	COOKIE	COOKIES	COOKING	COOK	COOKY
COOLER	COOL	COOP	COO	COPPER	COPY
CORD	CORK	CORNER	CORN	CORRECT	COST
COT	COTTAGE	COTTON	COUCH	COUGH	COULDN T
COULD	COUNTER	COUNTRY	COUNT	COUNTY	COURSE
COURT	COUSIN	COVER	COWARDLY	COWARD	COWBOY
COW	COZY	CRAB	CRACKER	CRACK	CRADLE
CRAMPS	CRANBERRY	CRANK	CRANKY	CRASH	CRAWL
CRAZY	CREAM	CREAMY	CREEK	CREEP	CREPT
CRIED	CRIES	CROAK	CROOKED	CROOK	CROP
CROSSING	CROSS-EYED	CROSS	CROWDED	CROWD	CROWN
CROW	CRUEL	CRUMBLE	CRUMB	CRUSH	CRUST
CRY	CUB	CUFF	CUPBOARD	CUPFUL	CUP
CURE	CURL	CURLY	CURTAIN	CURVE	CUSHION
CUSTARD	CUSTOMER	CUTE	CUT	CUTTING	DAB
DADDY	DAD	DAILY	DAIRY	DAISY	DAMAGE
DAME	DAMP	DAM	DANCER	DANCE	DANCING
DANDY	DANGEROUS	DANGER	DARE	DARKNESS	DARLING
DARN	DART	DASH	DATE	DAUGHTER	DAWN
DAYBREAK	DAY	DAYTIME	DEAD	DEAF	DEAL
DEAR	DEATH	DECEMBER	DECIDE	DECK	DEED
DEEP	DEER	DEFEAT	DEFEND	DEFENSE	DELIGHT
DEN	DENTIST	DEPEND	DEPOSIT	DESCRIBE	DESERT
DESERVE	DESIRE	DESK	DESTROY	DEVIL	DEW
DIAMOND	DIDN T	DID	DIED	DIE	DIES
DIFFERENCE	DIFFERENT	DIG	DIME	DIM	DINE
DING-DONG	DINNER	DIP	DIRECTION	DIRECT	DIRT
DIRTY	DISCOVER	DISH	DISLIKE	DISMISS	DITCH

DIVER	DIVE	DIVIDE	DOCK	DOCTOR	DOESN T
DOES	DOG	DOLLAR	DOLL	DOLLY	DON T
DONE	DONKEY	DOORBELL	DOORKNOB	DOOR	DOORSTEP
DOPE	DO	DOT	DOUBLE	DOUGH	DOVE
DOWN	DOWNSTAIRS	DOWNTOWN	DOZEN	DRAG	DRAIN
DRANK	DRAWER	DRAWING	DRAW	DREAM	DRESSER
DRESSMAKER	DRESS	DREW	DRIED	DRIFT	DRILL
DRINK	DRIP	DRIVEN	DRIVER	DRIVE	DROP
DROVE	DROWN	DROWSY	DRUG	DRUM	DRUNK
DRY	DUCK	DUE	DUG	DULL	DUMB
DUMP	DURING	DUST	DUSTY	DUTY	DWARF
DWELL	DWELT	DYING	EACH	EAGER	EAGLE
EARLY	EARN	EAR	EARTH	EASTERN	EAST
EASY	EATEN	EAT	EDGE	EGG	EH
EIGHTEEN	EIGHTH	EIGHT	EIGHTY	EITHER	ELBOW
ELDER	ELDEST	ELECTRICITY	ELECTRIC	ELEPHANT	ELEVEN
ELF	ELM	ELSE	ELSEWHERE	EMPTY	ENDING
END	ENEMY	ENGINEER	ENGINE	ENGLISH	ENJOY
ENOUGH	ENTER	ENVELOPE	EQUAL	ERASER	ERASE
ERRAND	ESCAPE	EVENING	EVEN	EVER	EVERYBODY
EVERYDAY	EVERYONE	EVERY	EVERYTHING	EVERYWHERE	EVE
EVIL	EXACT	EXCEPT	EXCHANGE	EXCITED	EXCITING
EXCUSE	EXIT	EXPECT	EXPLAIN	EXTRA	EYEBROW
EYE	FABLE	FACE	FACING	FACTORY	FACT
FAIL	FAINT	FAIR	FAIRY	FAITH	FAKE
FALL	FALSE	FAMILY	FANCY	FAN	FARAWAY
FARE	FAR-OFF	FARMER	FARMING	FARM	FAR
FARTHER	FASHION	FASTEN	FAST	FATHER	FAT
FAULT	FAVORITE	FAVOR	FEAR	FEAST	FEATHER
FEBRUARY	FED	FEED	FEEL	FEET	FELLOW
FELL	FELT	FENCE	FEVER	FEW	FIB
FIDDLE	FIELD	FIFE	FIFTEEN	FIFTH	FIFTY
FIGHT	FIG	FIGURE	FILE	FILL	FILM
FINALLY	FIND	FINE	FINGER	FINISH	FIREARM
FIRECRACKER	FIREPLACE	FIRE	FIREWORKS	FIRING	FIRST
FISHERMAN	FISH	FIST	FIT	FITS	FIVE
FIX	FLAG	FLAKE	FLAME	FLAP	FLASHLIGHT
FLASH	FLAT	FLEA	FLESH	FLEW	FLIES
FLIGHT	FLIP-FLOP	FLIP	FLOAT	FLOCK	FLOOD
FLOOR	FLOP	FLOUR	FLOWER	FLOWERY	FLOW
FLUTTER	FLY	FOAM	FOGGY	FOG	FOLD
FOLKS	FOLLOWING	FOLLOW	FOND	FOOD	FOOLISH
FOOL	FOOTBALL	FOOTPRINT	FOOT	FOREHEAD	FOREST
FORGET	FORGIVE	FORGOT	FORGOTTEN	FORK	FORM
FOR	FORTH	FORTH	FORT	FORTUNE	FORTY
FORWARD	FOUGHT	FOUND	FOUNTAIN	FOUR	FOURTEEN
FOX	FRAME	FREEDOM	FREE	FREEZE	FREIGHT
FRENCH	FRESH	FRET	FRIDAY	FRIED	FRIENDLY
FRIEND	FRIENDSHIP	FRIGHTEN	FROG	FROM	FRONT
FROST	FROWN	FROZE	FRUIT	FRY	FUDGE
FUEL	FULL	FULLY	FUNNY	FUN	FURNITURE
FUR	FURTHER	FUZZY	GAIN	GALLON	GALLOP
GAME	GANG	GARAGE	GARBAGE	GARDEN	GASOLINE
GAS	GATE	GATHER	GAVE	GAY	GEAR
GEESE	GENERAL	GENTLEMAN	GENTLEMEN	GENTLE	GEOGRAPHY
GET	GETTING	GIANT	GIFT	GINGERBREAD	GIRL

GIVEN	GIVE	GIVING	GLADLY	GLAD	GLANCE
GLASSES	GLASS	GLEAM	GLIDE	GLORY	GLOVE
GLOW	GLUE	GOAL	GOAT	GOBBLE	GODG
GODMOTHER	GOD	GOES	GOING	GOLDEN	GOLDFISH
GOLD	GOLF	GONE	GOOD-BYE	GOOD-BY	GOOD-LOOKING
GOODNESS	GOOD	GOODS	GOODY	GOOSEBERRY	GOOSE
GO	GOT	GOVERNMENT	GOVERN	GOWN	GRAB
GRACIOUS	GRADE	GRAIN	GRANDCHILDRE	GRANDCHILD	GRANDDAUGHTE
GRANDFATHER	GRANDMA	GRANDMOTHER	GRANDPA	GRAND	GRANDSON
GRANDSTAND	GRAPEFRUIT	GRAPE	GRAPES	GRASSHOPPER	GRASS
GRATEFUL	GRAVEL	GRAVE	GRAVEYARD	GRAVY	GRAY
GRAZE	GREASE	GREAT	GREEN	GREET	GREW
GRIND	GROAN	GROCERY	GROUND	GROUP	GROVE
GROW	GUARD	GUESS	GUEST	GUIDE	GULF
GUM	GUNPOWDER	GUN	GUY	HABIT	HADN T
HAD	HAIL	HAIRCUT	HAIRPIN	HAIR	HALF
HALL	HALT	HAMMER	HAM	HANDFUL	HANDKERCHIEF
HANDLE	HAND	HANDWRITING	HANG	HAPPEN	HAPPILY
HAPPINESS	HAPPY	HARBOR	HARDLY	HARD	HARDSHIP
HARDWARE	HARE	HARK	HARM	HARNESS	HARP
HARVEST	HA	HASN T	HAS	HASTEN	HASTE
HASTY	HATCHET	HATCH	HATE	HAT	HAUL
HAVEN T	HAVE	HAVING	HAWK	HAYFIELD	HAY
HAYSTACK	HE D	HE LL	HE S	HEADACHE	HEAD
HEAL	HEALTH	HEALTHY	HEAP	HEARD	HEARING
HEAR	HEART	HEATER	HEAT	HEAVEN	HEAVY
HEEL	HEIGHT	HELD	HELLO	HELL	HELMET
HELPER	HELPFUL	HELP	HEM	HENHOUSE	HEN
HERD	HERE S	HERE	HERO	HER	HERSELF
HERS	HE	HEY	HICKORY	HIDDEN	HIDE
HID	HIGH	HIGHWAY	HILL	HILLSIDE	HILLTOP
HILLY	HIM	HIMSELF	HIND	HINT	HIP
HIRE	HIS	HISS	HISTORY	HITCH	HIT
HIVE	HOE	HOG	HOLDER	HOLD	HOLE
HOLIDAY	HOLLOW	HOLY	HOMELY	HOME	HOMESICK
HONEST	HONEYBEE	HONEYMOON	HONEY	HONK	HONOR
HOOD	HOOF	HOOK	HOOP	HOPEFUL	HOPELESS
HOPE	HOP	HORN	HORSEBACK	HORSE	HORSESHOE
HO	HOSE	HOSPITAL	HOST	HOTEL	HOT
HOUND	HOUR	HOUSE	HOUSETOP	HOUSEWIFE	HOUSEWORK
HOWEVER	HOWL	HOW	HUGE	HUG	HUMBLE
HUMP	HUM	HUNDRED	HUNGER	HUNGRY	HUNG
HUNK	HUNTER	HUNT	HURRAH	HURRIED	HURRY
HURT	HUSBAND	HUSH	HUT	HYMN	I D
I LL	I M	I VE	ICE	ICY	IDEAL
IDEA	IF	ILL	IMPORTANT	IMPOSSIBLE	IMPROVE
INCHES	INCH	INCOME	INDEED	INDIAN	INDOORS
INK	INN	IN	INSECT	INSIDE	INSTANT
INSTEAD	INSULT	INTEND	INTERESTED	INTERESTING	INTO
INVITE	IRON	I	ISLAND	ISN T	IS
IT S	IT	ITSELF	ITS	IVORY	IVY
JACKET	JACKS	JAIL	JAM	JANUARY	JAR
JAW	JAY	JELLYFISH	JELLY	JERK	JIG
JOB	JOCKEY	JOIN	JOKE	JOKING	JOLLY
JOURNEY	JOYFUL	JOYOUS	JOY	JUDGE	JUG
JUICE	JUICY	JULY	JUMP	JUNE	JUNIOR

JUNK	JUST	KEEN	KEEP	KEPT	KETTLE
KEY	KICK	KID	KILLED	KILL	KINDLY
KINDNESS	KIND	KINGDOM	KING	KISS	KITCHEN
KITE	KITTEN	KITTY	KNEEL	KNEE	KNEW
KNIFE	KNIT	KNIVES	KNOB	KNOCK	KNOT
KNOWN	KNOW	LACE	LADDER	LADIES	LAD
LADY	LAID	LAKE	LAMB	LAME	LAMP
LAND	LANE	LANGUAGE	LANTERN	LAP	LARD
LARGE	LASH	LASS	LAST	LATE	LAUGH
LAUNDRY	LAWN	LAW	LAWYER	LAY	LAZY
LEADER	LEAD	LEAF	LEAK	LEAN	LEAP
LEARNED	LEARN	LEAST	LEATHER	LEAVE	LEAVING
LED	LEFT	LEG	LEMONADE	LEMON	LEND
LENGTH	LESSON	LESS	LET S	LET	LETTER
LETTING	LETTUCE	LEVEL	LIBERTY	LIBRARY	LICE
LICK	LID	LIE	LIFE	LIFT	LIGHTNESS
LIGHTNING	LIGHT	LIKELY	LIKE	LIKING	LILY
LIMB	LIME	LIMP	LINEN	LINE	LION
LIP	LISTEN	LIST	LIT	LITTLE	LIVELY
LIVER	LIVE	LIVES	LIVING	LIZARD	LOAD
LOAF	LOAN	LOAVES	LOCK	LOCOMOTIVE	LOG
LONELY	LONE	LONESOME	LONG	LOOKOUT	LOOK
LOOP	LOOSE	LORD	LOSER	LOSE	LOSS
LOST	LOT	LOUD	LOVELY	LOVER	LOVE
LOW	LUCK	LUCKY	LUMBER	LUMP	LUNCH
LYING	MACHINERY	MACHINE	MADE	MAD	MAGAZINE
MAGIC	MAID	MAILBOX	MAILMAN	MAIL	MAJOR
MAKE	MAKING	MALE	MAMA	MAMMA	MANAGER
MANE	MANGER	MAN	MANY	MAPLE	MAP
MARBLE	MARCH	MARE	MARKET	MARK	MARRIAGE
MARRIED	MARRY	MA	MASK	MASTER	MAST
MATCH	MAT	MATTER	MATTRESS	MAYBE	MAYOR
MAYPOLE	MAY	MEADOW	MEAL	MEAN	MEANS
MEANT	MEASURE	MEAT	MEDICINE	MEETING	MEET
MELT	MEMBER	MEND	MEN	MEOW	MERRY
ME	MESSAGE	MESS	METAL	MET	MEW
MICE	MIDDLE	MIDNIGHT	MIGHT	MIGHTY	MILE
MILKMAN	MILK	MILLER	MILLION	MILL	MIND
MINER	MINE	MINT	MINUTE	MIRROR	MISCHIEF
MISSPELL	MISS	MISTAKE	MISTY	MITTEN	MITT
MIX	MOMENT	MONDAY	MONEY	MONKEY	MONTH
MOONLIGHT	MOON	MOO	MOOSE	MOP	MORE
MORNING	MORROW	MOSS	MOSTLY	MOST	MOTHER
MOTOR	MOUNTAIN	MOUNT	MOUSE	MOUTH	MOVE
MOVIE	MOVIES	MOVING	MOW	MR.	MRS.
MUCH	MUDDY	MUD	MUG	MULE	MULTIPLY
MURDER	MUSIC	MUST	MY	MYSELF	NAIL
NAME	NAPKIN	NAP	NARROW	NASTY	NAUGHTY
NAVY	NEARBY	NEARLY	NEAR	NEAT	NECK
NECKTIE	NEEDLE	NEEDN T	NEED	NEGRO	NEIGHBORHOOD
NEIGHBOR	NEITHER	NERVE	NEST	NET	NEVERMORE
NEVER	NEW	NEWSPAPER	NEWS	NEXT	NIBBLE
NICE	NICKEL	NIGHTGOWN	NIGHT	NINE	NINETEEN
NINETY	NOBODY	NOD	NOISE	NOISY	NONE
NOON	NOR	NORTHERN	NORTH	NO	NOSE
NOTE	NOTHING	NOTICE	NOT	NOVEMBER	NOWHERE

NOW	NUMBER	NURSE	NUT	OAK	OAR
OATMEAL	OATS	OBEY	OCEAN	OCLOCK	OCTOBER
ODD	OFFER	OFFICER	OFFICE	OFF	OF
OFTEN	OH	OIL	OLD-FASHIONE	OLD	ONCE
ONE	ONION	ONLY	ON	ONWARD	OPEN
ORANGE	ORCHARD	ORDER	ORE	ORGAN	OR
OTHER	OTHERWISE	OUCH	OUGHT	OUR	OURSELVES
OURS	OUTDOORS	OUTFIT	OUTLAW	OUTLINE	OUT
OUTSIDE	OUTWARD	OVEN	OVERALLS	OVERCOAT	OVEREAT
OVERHEAD	OVERHEAR	OVERNIGHT	OVER	OVERTURN	OWE
OWING	OWL	OWNER	OWN	OX	PACE
PACKAGE	PACK	PAD	PAGE	PAID	PAIL
PAINFUL	PAIN	PAINTER	PAINTING	PAINT	PAIR
PALACE	PALE	PAL	PAMERICA	PANCAKE	PANE
PAN	PANSY	PANTS	PAPA	PAPER	PARADE
PARDON	PARENT	PARK	PARTLY	PARTNER	PART
PARTY	PA	PASSENGER	PASS	PASTE	PAST
PASTURE	PATCH	PATH	PAT	PATTER	PAVEMENT
PAVE	PAW	PAYMENT	PAY	PEACEFUL	PEACE
PEACHES	PEACH	PEAK	PEANUT	PEARL	PEAR
PEA	PEAS	PECK	PEEK	PEEL	PEEP
PEG	PENCIL	PENNY	PEN	PEOPLE	PEPPERMINT
PEPPER	PERFUME	PERHAPS	PERSON	PET	PHONE
PIANO	PICKLE	PICK	PICNIC	PICTURE	PIECE
PIE	PIGEON	PIGGY	PIG	PILE	PILLOW
PILL	PINEAPPLE	PINE	PINK	PIN	PINT
PIPE	PISTOL	PITCHER	PITCH	PIT	PITY
PLACE	PLAIN	PLANE	PLAN	PLANT	PLATE
PLATFORM	PLATTER	PLAYER	PLAYGROUND	PLAYHOUSE	PLAYMATE
PLAY	PLAYTHING	PLEASANT	PLEASE	PLEASURE	PLENTY
PLOW	PLUG	PLUM	POCKETBOOK	POCKET	POEM
POINT	POISON	POKE	POLE	POLICEMAN	POLICE
POLISH	POLITE	POND	PONIES	PONY	POOL
POOR	POPCORN	POPPED	POP	PORCH	PORK
POSSIBLE	POSTAGE	POSTMAN	POST	POTATOES	POTATO
POT	POUND	POUR	POWDER	POWERFUL	POWER
PRAISE	PRAYER	PRAY	PREPARE	PRESENT	PRETTY
PRICE	PRICK	PRINCE	PRINCESS	PRINT	PRISON
PRIZE	PROMISE	PROPER	PROTECT	PROUD	PROVE
PRUNE	PUBLIC	PUDDLE	PUFF	PULL	PUMPKIN
PUMP	PUNCH	PUNISH	PUPIL	PUPPY	PUP
PURE	PURPLE	PURSE	PUSH	PUSS	PUSSYCAT
PUSSY	PUT	PUTTING	PUZZLE	QUACK	QUARTER
QUART	QUEEN	QUEER	QUESTION	QUICKLY	QUICK
QUIET	QUILT	QUITE	QUIT	RABBIT	RACE
RACK	RADIO	RADISH	RAG	RAILROAD	RAIL
RAILWAY	RAINBOW	RAIN	RAINY	RAISE	RAISIN
RAKE	RAM	RANCH	RANG	RAN	RAPIDLY
RAP	RATE	RATHER	RAT	RATTLE	RAW
RAY	REACH	READER	READING	READ	READY
REALLY	REAL	REAP	REAR	REASON	REBUILD
RECEIVE	RECESS	RECORD	REDBIRD	REDBREAST	RED
REFUSE	REINDEER	REJOICE	REMAIN	REMEMBER	REMIND
REMOVE	RENT	REPAIR	REPAY	REPEAT	REPORT
REST	RETURN	REVIEW	REWARD	RIBBON	RIB
RICE	RICH	RIDDLE	RIDER	RIDE	RIDING

APPENDIX A (Continued)

RID	RIGHT	RIM	RING	RIPE	RIP
RISE	RISING	RIVER	ROAD	ROADSIDE	ROAR
ROAST	ROBBER	ROBE	ROBIN	ROB	ROCKET
ROCK	ROCKY	RODE	ROLLER	ROLL	ROOF
ROOM	ROOSTER	ROOT	ROPE	ROSEBUD	ROSE
ROT	ROTTEN	ROUGH	ROUND	ROUTE	ROWBOAT
ROW	ROYAL	RUBBED	RUBBER	RUBBISH	RUB
RUG	RULER	RULE	RUMBLE	RUNG	RUNNER
RUNNING	RUN	RUSH	RUST	RUSTY	RYE
SACK	SADDLE	SADNESS	SAD	SAFE	SAFETY
SAID	SAILBOAT	SAILOR	SAIL	SAINT	SALAD
SALE	SALT	SAME	SAND	SANDWICH	SANDY
SANG	SANK	SAP	SASH	SATIN	SATISFACTORY
SAT	SATURDAY	SAUSAGE	SAVAGE	SAVE	SAVINGS
SAW	SAY	SCAB	SCALES	SCARE	SCARF
SCHOOLBOY	SCHOOLHOUSE	SCHOOLMASTER	SCHOOLROOM	SCHOOL	SCORCH
SCORE	SCRAPE	SCRAP	SCRATCH	SCREAM	SCREEN
SCREW	SCRUB	SEAL	SEAM	SEARCH	SEA
SEASON	SEAT	SECOND	SECRET	SEED	SEEING
SEEK	SEEM	SEEN	SEE	SEESAW	SELECT
SELFISH	SELF	SELL	SEND	SENSE	SENTENCE
SENT	SEPARATE	SEPTEMBER	SERVANT	SERVE	SERVICE
SET	SETTING	SETTLEMENT	SETTLE	SEVEN	SEVENTEEN
SEVENTH	SEVENTY	SEVERAL	SEW	SHADE	SHADOW
SHADY	SHAKER	SHAKE	SHAKING	SHALL	SHAME
SHAN T	SHAPE	SHARE	SHARP	SHAVE	SHE D
SHE LL	SHE S	SHEAR	SHEARS	SHED	SHEEP
SHEET	SHELF	SHELL	SHEPHERD	SHE	SHINE
SHINING	SHINY	SHIP	SHIRT	SHOCK	SHOEMAKER
SHOE	SHONE	SHOOK	SHOOT	SHOPPING	SHOP
SHORE	SHORT	SHOT	SHOULDER	SHOULDN T	SHOULD
SHOUT	SHOVEL	SHOWER	SHOW	SHUT	SHY
SICKNESS	SICK	SIDE	SIDEWALK	SIDEWAYS	SIGH
SIGHT	SIGN	SILENCE	SILENT	SILK	SILL
SILLY	SILVER	SIMPLE	SINCE	SINGER	SINGLE
SING	SINK	SIN	SIP	SIR	SIS
SISSY	SISTER	SIT	SITTING	SIX	SIXTEEN
SIXTH	SIXTY	SIZE	SKATER	SKATE	SKIN
SKIP	SKIRT	SKI	SKY	SLAM	SLAP
SLATE	SLAVE	SLED	SLEEP	SLEEPY	SLEEVE
SLEIGH	SLEPT	SLICE	SLIDE	SLID	SLING
SLIPPED	SLIPPER	SLIPPERY	SLIP	SLIT	SLOWLY
SLOW	SLY	SMACK	SMALL	SMART	SMELL
SMILE	SMOKE	SMOOTH	SNAIL	SNAKE	SNAPPING
SNAP	SNEEZE	SNOWBALL	SNOWFLAKE	SNOW	SNOWY
SNUFF	SNUG	SOAK	SOAP	SOB	SOCKS
SODA	SOD	SOFA	SOFT	SOIL	SOLDIER
SOLD	SOLE	SOMEBODY	SOMEHOW	SOMEONE	SOME
SOMETHING	SOMETIME	SOMETIMES	SOMEWHERE	SONG	SON
SOON	SORE	SORROW	SORRY	SORT	SO
SOUL	SOUND	SOUP	SOUR	SOUTHERN	SOUTH
SPACE	SPADE	SPANK	SPARROW	SPEAKER	SPEAK
SPEAR	SPEECH	SPEED	SPELLING	SPELL	SPEND
SPENT	SPIDER	SPIKE	SPILL	SPINACH	SPIN
SPIRIT	SPIT	SPLASH	SPOIL	SPOKE	SPOOK
SPOON	SPORT	SPOT	SPREAD	SPRING	SPRINGTIME

SPRINKLE	SQUARE	SQUASH	SQUEAK	SQUEEZE	SQUIRREL
STABLE	STACK	STAGE	STAIR	STALL	STAMP
STAND	STARE	STAR	START	STARVE	STATE
STATION	STAY	STEAK	STEAL	STEAMBOAT	STEAMER
STEAM	STEEL	STEEPLE	STEEP	STEER	STEM
STEPPING	STEP	STICK	STICKY	STIFF	STILLNESS
STILL	STING	STIR	STITCH	STOCKING	STOCK
STOLE	STONE	STOOD	STOOL	STOOP	STOPPED
STOPPING	STOP	STORE	STORIES	STORK	STORM
STORMY	STORY	STOVE	STRAIGHT	STRANGER	STRANGE
STRAP	STRAWBERRY	STRAW	STREAM	STREET	STRETCH
STRING	STRIPES	STRIP	STRONG	STUCK	STUDY
STUFF	STUMP	STUNG	SUBJECT	SUCH	SUCK
SUDDEN	SUFFER	SUGAR	SUIT	SUMMER	SUM
SUNDAY	SUNFLOWER	SUNG	SUNK	SUNLIGHT	SUNNY
SUNRISE	SUN	SUNSET	SUNSHINE	SUPPER	SUPPOSE
SURELY	SURE	SURFACE	SURPRISE	SWALLOW	SWAMP
SWAM	SWAN	SWAT	SWEAR	SWEATER	SWEAT
SWEEP	SWEETHEART	SWEETNESS	SWEET	SWELL	SWEPT
SWIFT	SWIMMING	SWIM	SWING	SWITCH	SWORD
SWORE	TABLECLOTH	TABLE	TABLESPOON	TABLET	TACK
TAG	TAILOR	TAIL	TAKEN	TAKE	TAKING
TALE	TALKER	TALK	TALL	TAME	TANK
TAN	TAPE	TAP	TARDY	TAR	TASK
TASTE	TAUGHT	TAX	TEACHER	TEACH	TEAM
TEAR	TEA	TEASE	TEASPOON	TEETH	TELEPHONE
TELL	TEMPER	TENNIS	TEN	TENT	TERM
TERRIBLE	TEST	THANKFUL	THANK	THANKSGIVING	THANKS
THAN	THAT S	THAT	THEATER	THEE	THEIR
THEM	THEN	THERE	THE	THESE	THEY D
THEY LL	THEY RE	THEY VE	THEY	THICK	THIEF
THIMBLE	THING	THINK	THIN	THIRD	THIRSTY
THIRTEEN	THIRTY	THIS	THORN	THO	THOSE
THOUGH	THOUGHT	THOUSAND	THREAD	THREE	THREW
THROAT	THRONE	THROUGH	THROWN	THROW	THUMB
THUNDER	THURSDAY	THY	TICKET	TICKLE	TICK
TIE	TIGER	TIGHT	TILL	TIME	TINKLE
TIN	TINY	TIP	TIPTOE	TIRED	TIRE
TITLE	TOAD	TOADSTOOL	TOAST	TOBACCO	TODAY
TOE	TOGETHER	TOILET	TOLD	TOMATO	TOMORROW
TONE	TONGUE	TONIGHT	TON	TOOK	TOOL
TOO	TOOTHBRUSH	TOOTHPICK	TOOTH	TOOT	TOP
TORE	TORN	TO	TOSS	TOUCH	TOWARD S
TOWARD	TOWEL	TOWER	TOWN	TOW	TOY
TRACE	TRACK	TRADE	TRAIN	TRAMP	TRAP
TRAY	TREASURE	TREAT	TREE	TRICK	TRICYCLE
TRIED	TRIM	TRIP	TROLLEY	TROUBLE	TRUCK
TRUE	TRULY	TRUNK	TRUST	TRUTH	TRY
TUB	TUESDAY	TUG	TULIP	TUMBLE	TUNE
TUNNEL	TURKEY	TURN	TURTLE	TWELVE	TWENTY
TWICE	TWIG	TWIN	TWO	UGLY	UMBRELLA
UNCLE	UNDER	UNDERSTAND	UNDERWEAR	UNDRESS	UNFAIR
UNFINISHED	UNFOLD	UNFRIENDLY	UNHAPPY	UNHURT	UNIFORM
UNITED STATE	UNKIND	UNKNOWN	UNLESS	UNPLEASANT	UNTIL
UNWILLING	UPON	UPPER	UP	UPSET	UPSIDE
UPSTAIRS	UPTOWN	UPWARD	USED	USEFUL	USE

US	VALENTINE	VALLEY	VALUABLE	VALUE	VASE
VEGETABLE	VELVET	VERY	VESSEL	VICTORY	VIEW
VILLAGE	VINE	VIOLET	VISITOR	VISIT	VOICE
VOTE	WAGON	WAG	WAIST	WAIT	WAKEN
WAKE	WALK	WALL	WALNUT	WANT	WARM
WARN	WAR	WASHER	WASH	WASHTUB	WASN T
WAS	WASTE	WATCHMAN	WATCH	WATERMELON	WATERPROOF
WATER	WAVE	WAX	WAY	WAYSIDE	WE D
WE LL	WE RE	WE VE	WEAKEN	WEAKNESS	WEAK
WEALTH	WEAPON	WEAR	WEARY	WEATHER	WEAVE
WEB	WEDDING	WEDNESDAY	WEED	WEEK	WEEP
WEE	WEIGH	WELCOME	WELL	WENT	WERE
WE	WESTERN	WEST	WET	WHALE	WHAT S
WHAT	WHEAT	WHEEL	WHENEVER	WHEN	WHERE
WHICH	WHILE	WHIPPED	WHIP	WHIRL	WHISKY
WHISPER	WHISTLE	WHITE	WHO D	WHO LL	WHO S
WHOLE	WHOM	WHO	WHOSE	WHY	WICKED
WIDE	WIFE	WIGGLE	WILDCAT	WILD	WILLING
WILLOW	WILL	WINDMILL	WINDOW	WIND	WINDY
WINE	WING	WINK	WINNER	WIN	WINTER
WIPE	WIRE	WISE	WISH	WITCH	WITHOUT
WITH	WIT	WOKE	WOLF	WOMAN	WOMEN
WON T	WONDERFUL	WONDER	WON	WOODEN	WOODPECKER
WOOD	WOODS	WOOLEN	WOOL	WORD	WORE
WORKER	WORKMAN	WORK	WORLD	WORM	WORN
WORRY	WORSE	WORST	WORTH	WOULDN T	WOULD
WOUND	WOVE	WRAPPED	WRAP	WRECK	WREN
WRING	WRITE	WRITING	WRITTEN	WRONG	WROTE
WRUNG	YARD	YARN	YEAR	YELLOW	YELL
YES	YESTERDAY	YET	YOLK	YONDER	YOU D
YOURSELF	YOURSELVES	YOURS	YOU	YOUTH	ZZZZZZZZZZZZ
ZZZZZZZZZZZZ	ZZZZZZZZZZZZ	ZZZZZZZZZZZZ	ZZZZZZZZZZZZ	ZZZZZZZZZZZZ	ZZZZZZZZZZZZ
ZZZZZZZZZZZZ	ZZZZZZZZZZZZ	ZZZZZZZZZZZZ	ZZZZZZZZZZZZ	ZZZZZZZZZZZZ	ZZZZZZZZZZZZ
ZZZZZZZZZZZZ	ZZZZZZZZZZZZ	ZZZZZZZZZZZZ	ZZZZZZZZZZZZ	ZZZZZZZZZZZZ	ZZZZZZZZZZZZ
ZZZZZZZZZZZZ	ZZZZZZZZZZZZ	ZZZZZZZZZZZZ	ZZZZZZZZZZZZ	ZZZZZZZZZZZZ	ZZZZZZZZZZZZ
ZZZZZZZZZZZZ	ZZZZZZZZZZZZ	ZZZZZZZZZZZZ	ZZZZZZZZZZZZ	ZZZZZZZZZZZZ	ZZZZZZZZZZZZ
ZZZZZZZZZZZZ	ZZZZZZZZZZZZ	ZZZZZZZZZZZZ	ZZZZZZZZZZZZ	ZZZZZZZZZZZZ	ZZZZZZZZZZZZ
ZZZZZZZZZZZZ	ZZZZZZZZZZZZ	ZZZZZZZZZZZZ	ZZZZZZZZZZZZ	ZZZZZZZZZZZZ	ZZZZZZZZZZZZ
ZZZZZZZZZZZZ	ZZZZZZZZZZZZ	ZZZZZZZZZZZZ	ZZZZZZZZZZZZ	ZZZZZZZZZZZZ	ZZZZZZZZZZZZ
ZZZZZZZZZZZZ	ZZZZZZZZZZZZ	ZZZZZZZZZZZZ	ZZZZZZZZZZZZ	ZZZZZZZZZZZZ	ZZZZZZZZZZZZ
ZZZZZZZZZZZZ	ZZZZZZZZZZZZ	ABSENSE	ABSOLUTLY	ABUNDENCE	ABUNDENT
ACADAMY	ACCELLERATOR	ACCEPTENCE	ACCESSABLE	ACCIDENTLY	ACCOMODATE
ACCOMODATION	ACCROSS	ACEDEMIC	ACHEIVEMENT	ACHIEVMENT	ACKNOWLEGE
ACKNOWLEGING	ACTUALY	ACUMULATE	ACUMULATING	ACURACY	ACURACY
ACURATE	ADEQUETE	ADMITTENCE	ADOLECENT	ADOLESENT	ADVANTAGOUS
ADVERTISMENT	AFAIR	AGGRESSIVE	AGGRIVATE	AGGRIVATING	AGRAVATE
AGRAVATING	AGRESIVE	AGRESSIVE	AGRIVATE	AGRIVATING	ALLEDGE
ALLEDGING	ALLEGIENCE	ALLOTED	ALLWAYS	ALOT	ALOTTED
ALPHEBET	ALRIGHT	AMOUNG	AMUNG	APARATUS	APINION
APOLIGIZE	APOLOJIZE	APPARANT	APPARANT	APPEARENCE	AQUAINT
AQUIRE	AQUIRING	ARGUEING	ARGUEMENT	ARGUEMENT	ARRANGMENT
ARTICAL	ARTIC	ASSASIN	ATHELETE	ATHELETIC	ATITUDE
ATTATCH	ATTENDENCE	ATTENDENT	AUDIANCE	AUGEST	AUTHORATATIV
AUTHORATY	AUTOES	AUTOES	BACKROUND	BALENCE	BALENCING
BALOON	BARBEROUS	BARGIN	BASICLY	BASICLY	BEATIFUL
BECOMMING	BEFOR	BEGGER	BEGINING	BELEIF	BELEIVE

APPENDIX A (Continued)

BELEIVING	BENIFICIAL	BENIFITED	BENIFITTED	BEUTIFUL	BEUTY
BIGER	BISCIT	BISCUT	BOOKEEPER	BOUNDRY	BRECKFAST
BRETH	BRITIN	BULETTIN	BULLETTIN	BURGLER	BURIEL
BURYED	BUSNESS	CAFATERIA	CALENDER	CAMOFLAGE	CAMPAIN
CAPTIN	CARFUL	CATAGORIES	CATAGORIES	CATAGORY	CATAGORY
CELLER	CEMETARIES	CEMETARY	CENTERY	CERTIN	CHALLANGE
CHANGABLE	CHARACTOR	CHARICTER	CHILDERN	CIELING	CIGARETE
CIGERETTE	CIGGARETE	CIGGARETTE	COFEE	COLLEDGE	COLLOSAL
COLLOSSAL	COLOSAL	COMERCIAL	COMITTEE	COMMING	COMMITEE
COMPARITIVE	COMPATABLE	COMPETANT	COMPLEAT	COMPLETLY	COMUNIST
CONCIEVABLE	CONCIEVE	CONCIOUS	CONSEAL	CONSENTRATE	CONSERN
CONSERT	CONSIENCE	CONSIOUS	CONSISTANCY	CONSISTANT	CONSISTANT
CONSTENT	CONTEMPERARY	CONTEMPORERY	CONTRAVERSY	CONTROLING	CONVIENCE
CONVINIENCE	CORELATE	CORIDOR	COROBORATE	COROBORATING	CORPERATION
CORRESPONDAN	COUNCEL	COURAGOUS	CRITICICM	CRITICISE	CRITISISM
CRUELY	CURICULAR	CURICULLAR	CURICULLUM	CURICULUM	CURRING
CUSTUM	DECENT	DEFERENCE	DEFINATE	DEMOCRASY	DEPENDANT
DESIDE	DESIREABLE	DESPARE	DEVELOPE	DICEASE	DICIPLE
DICIPLINE	DICIPLINING	DIFERENT	DIFFERANT	DIFFRENT	DIFICULT
DILEMA	DILIGANT	DILLEMA	DILLIGENT	DINNING	DISAPOINT
DISASTEROUS	DISATISFIED	DISCRIBE	DISCRIMANATE	DISCRIMANATI	DISCRIMANATI
DISCRIPTION	DISGISE	DISIPLE	DISIPLINE	DISIPLINING	DISPAIR
DISSAPOINT	DISSAPPOINT	DISTRUCTION	DOCTER	DOMINENT	DOMINENT
DONKIES	DROPED	ECHOS	ECSTACY	EFFICIENTCY	EIGTH
ELABERATE	ELECTRICTY	ELECTRISITY	ELIGABLE	ELIGABLE	EMBARASS
EMBARRAS	EMINANT	EMPERER	ENDEAVER	ENTERPRIZE	ENTIRLY
ENTRENCE	ENVIREMENT	ENVOLVE	EPADEMIC	EPEDEMIC	EQUIPED
EQUIPPMENT	ERATIC	EXAGERATE	EXAGERATING	EXAUST	EXCEDE
EXCEDING	EXCELLANCE	EXCELLANT	EXELLENCE	EXELLENT	EXEPT
EXERCIZE	EXERSISE	EXERSIZE	EXIBIT	EXISTANCE	EXITABLE
EXITE	EXITING	EXPENCE	EXPERAMENT	EXPEREMENT	EXPERIANCE
EXPLAINATION	EXPLINATION	EXTRACURICUL	EXTRACURRICU	EXTREAM	EXTREMLY
FACINATE	FACINATING	FACINATION	FALLICIES	FALLICY	FAMILAR
FAMILIER	FANTACY	FASINATE	FASINATING	FASINATION	FAVORIT
FEBUARY	FEILD	FICTICIOUS	FICTIOUS	FINALY	FORIEGN
FOURTY	FREIND	FRIENDLYNESS	FRIGHTNING	FUEDAL	FULLFIL
FUNDEMENTAL	GARENTEE	GAYETY	GENERALY	GILTY	GODESS
GON	GOVENOR	GOVERMENT	GRAMMER	GRANDURE	GRANDUR
GRUSOME	GUAGE	GUARENTEE	GUIDENCE	GYMNAZIUM	HANDKERCHEIF
HANKERCHIEF	HAPPENNED	HAPPYNESS	HARRAS	HARRASS	HEREDITERY
HEROS	HINDRENCE	HORRABLE	HORRABLY	HUMEROUS	HUMER
HUNDERD	HUNGERY	HURRIDLY	HYGEINE	HYPOCRACY	IDEALY
IGNORENCE	IGNORENT	IMAGINERY	IMEDIATE	IMENSE	IMFORMATION
IMMAGRANT	IMMEDIATLY	IMMENCE	IMMIGRENT	IMPERTINANT	IMPORTENCE
IMPORTENCE	IMPORTENT	IMPORTENT	INCONVIENCE	INCONVINIENC	INCREDABLE
INDEPENDANCE	INDEPENDANT	INDREDIANT	INEVITIBLE	INEVITIBLY	INFLUENCIAL
INGENOUS	INITATIVE	INOCENT	INTELECT	INTERFERANCE	INTERPERTATI
INTREST	IRRELEVENT	IRRESISTABLE	JEALUS	JELOUS	JEWELERY
JEWLERY	JOURNIES	JUVENIL	JUVIMILE	LABEROR	LABRATORIES
LABRATORY	LAYED	LEIZURE	LICENCE	LIESURE	LIKLIHOOD
LITRATURE	LIVLIEST	LIVLIHOOD	LONLEYNESS	LONLINESS	LOOSES
LUXERIES	LUXERY	MABE	MAGIZINE	MAGNIFICENSE	MAGNIFISENCE
MAINTAINANCE	MALICOUS	MANER	MANOUVER	MARIAGE	MARRAGE
MARRIDGE	MATERIEL	MATHAMATIC	MEDECAL	MEDECINE	MEDOW
MELENCHOLY	MENT	METEPHOR	MINAMUM	MINITURE	MISCHEIVOUS
MISPELL	MONKIES	MONOTONUS	MORELLY	MORGAGE	MORILLY

MOSQUITOS	MUSLE	MUSSLE	NARATIVE	NATURALY	NECCESARILY
NECCESITY	NECCESSARILY	NECCESSARY	NECCESSITY	NECESARY	NEGHBOR
NEGROS	NEICE	NEIGBOR	NESECITY	NESESARILY	NESESITY
NESESSARY	NESESSITY	NICKLE	NINETH	NOTICABLE	NUMERUS
NUMOROUS	NUSANCE	OBEDIANCE	OBEDIANT	OBSTICLE	OBSTICLE
OCASSION	OCCASSION	OCCURANCE	OCCURANCE	OCCURENCE	OCCURENCE
OCCURING	OCCURRANCE	OCCURRANCE	OMITED	OMMIT	OMMIT
OMMITTED	OPION	OPORTUNITY	OPPERATE	OPPONANT	OPTOMISM
OPTOMIST	ORGINIZE	ORIGINEL	ORIGINIL	PAMFLET	PARALIZE
PARIDISE	PARLAMENT	PARLEMENT	PARRALEL	PARRALLEL	PARTICULER
PARTICULER	PAYED	PEASENT	PECULIER	PERCIEVE	PERFORMENCE
PERMANANT	PERMENENT	PERSISTANT	PERSONALY	PERSONEL	PERSUE
PERSUING	PERTINANT	PHAMFLET	PHENOMINA	PHENOMINON	PLAGERISM
PLAGERIZE	PLAJERISM	PLAJERIZE	PLAUSABLE	PLAYWRITE	PLEASENT
POLITITIAN	POLITITION	PORTRATE	PORTRAT	PORTRIT	POSESSIVE
POSSESIVE	POTATOS	PRACTICLY	PRARIE	PRECEED	PRECICELY
PREEMIUM	PREFERED	PREFERING	PREFORMANCE	PREPERATION	PRESENSE
PRESTEGE	PREVELANT	PREVELENT	PRIMATIVE	PRIMATIVE	PRISINER
PRIVILEDGE	PROBIBLY	PROCEEDURE	PROFFESION	PROFFESSION	PROFFESSOR
PROMINANT	PROOVE	PURSUADE	QUANDITY	QUESTIONAIRE	QUIZES
RADIOES	RAIDO	RECCOMEND	RECEVE	RECEVING	RECIEVE
RECIEVING	RECOMEND	RECONIZE	REFERANCE	REFERED	REFERING
REHERSAL	RELEGION	RELEIVE	RELEIVING	RELETIVE	RELEVENT
RELIGEON	RELITIVE	REMBER	REMEMBERANCE	REPITITION	REPITITION
REPITITION	REPRESENTITI	RESPONCE	RESTARANT	RESTARONT	RESTURANT
RESTURONT	REVEEL	RIDACULE	RIDECULE	RIDECULING	RIGHTOUS
ROOMATE	RYTHM	SAFTY	SATERDAY	SAYED	SECRETERIES
SECRETERY	SEIGE	SENCE	SENE	SENTANCE	SEPERATE
SEPERATING	SEPERATION	SERVENT	SEVEREL	SHEPERD	SHERRIFF
SHERRIF	SHINNING	SHINNY	SIEZE	SIEZING	SIGNIFICENCE
SIGNITURE	SIMILER	SIMPEL	SINCERLY	SOCIATY	SOPHMORE
SOUVINIR	SPAGETTI	SPEACH	SPONSER	STEPED	STOMACK
STOPED	STOPING	STORYS	STRECH	STRENTH	STUBORN
STUDING	SUBTILE	SUBUB	SUCEED	SUCESS	SUGER
SUGEST	SUMARIES	SUMARY	SUPERINTENDA	SUPORT	SUPOSE
SUPRESS	SUPRISE	SUPRISING	SURBURB	SURGON	SURGURY
SURJERY	SURJON	SUROUND	SURPRIZE	SURPRIZING	SUSPENCE
SUTLE	SUVENIR	SWIMING	SWIMING	SYMBEL	SYMBLE
SYMBOLE	TALANT	TECNIQUE	TEMPERERY	TEMPERMENT	TEMPERTURE
TEMPORERILY	TEMPORERY	TENDANCY	TENDANCY	TERRABLE	TERRABLY
THANKYOU	THERFORE	THIER	THOUSEND	THROUGHLY	TOLERENCE
TOLERENT	TOMATOS	TOMMOROW	TOMMORROW	TRAGADIES	TRAGADY
TRAJEDIES	TRAJEDY	TRANSFERED	TREASUROR	TRESURER	TRUELY
TURKIES	TYRANY	UNATURAL	UNECESSARY	UNTILL	USEING
USFUL	USLESS	VACCUM	VACCUUM	VALUBLE	VARIOS
VARIUS	VEGATABLE	VEGTABLE	VEIW	VENGENCE	VILLIN
WENODSY	WENEDSDAY	WETHER	WICH	WIEGHT	WIERD
WRITEN	WRITTING	YEILD	HOW	WHAT	WHERE
WHICH	WHOM	WHO	WHOSE	ZZZZZZZZZZZZ	ZZZZZZZZZZZZ
ZZZZZZZZZZZZ	AFRICAN	AFRICA	ALL	ANOTHER	AN
ANYBODY	ANYONE	ANY	A	BLOOD	BOTH
CAPTAIN	CHINA	CHINESE	CIVILIZATION	COLLEGE	COLONEL
CORPORAL	COURAGE	CROME	DAVID	DR	EACH
EIGHT	EITHER	ENGLAND	ENGLISH	EVERYBODY	EVERYONE
EVERY	FEW	FIVE	FOOD	FOUR	FRANCE
FRENCH	GENERAL	GEORGE	GERMAN	GERMANY	GOVERNOR

```
HALF          HER           HE            HIS           HONESTY       HUMAN
I             ITALIAN       ITALY         IT            ITS           JAMES
JANE          JEAN          JOHN          JUSTICE       LESS          LIEUTENANT
LITTLE        MAJOR         MANKIND       MAN           MANY          MARY
MAYOR         MEN           MICHAEL       MINE          MORE          MOST
MR            MUCH          MY            NEITHER       NINE          NOTHING
ONE           OTHER         PAUL          PEACE         PEOPLE        PETER
PLENTY        PREMIER       PRESIDENT     PRIVATE       PROFESSOR     ROBERT
RUSSIAN       RUSSIA        SCHOOL        SERGEANT      SEVEN         SEVERAL
SHE           SIR           SIX           SOCIETY       SOMEBODY      SOMEONE
SOME          SPAIN         SPANISH       SUCH          SUE           SURVIVAL
TELEVISION    TEN           THAT          THEIR         THE           THESE
THEY          THIS          THOSE         THREE         TIME          TWO
WAR           WATER         WE            WHOLE         WILLIAM       WOMEN
YOUR          YOU           ZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZ
ZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZ
ZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZ
ZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZ
ZZZZZZZZZZZZ
ACROSS
AS
HAS
IS
LESS
THUS
UNLESS
WAS
YES
ZZZZZZZZZZZZ
99191312331110103413171820211615191499999999999999999999999999999999
99999
$IBSYS
```

APPENDIX B

TABLE IV-11 (A)

Predictor $\underline{n}$ = 25

SHRUNKEN MULTIPLE-REGRESSION COEFFICIENTS

COMPUTED FROM WHERRY FORMULA

(See Chapter IV)

Discovered MULTR	Sample Size 100	125	150	175	200	225	250	275	300
.50	00	25	31	35	38	39	41	42	43
.51	10	27	33	37	39	41	42	43	44
.52	15	29	35	38	41	42	43	44	45
.53	19	32	37	40	42	44	45	46	46
.54	23	34	39	42	44	45	46	47	48
.55	26	36	40	43	45	46	47	48	49
.56	29	37	42	45	46	48	49	49	50
.57	31	39	43	46	48	49	50	51	51
.58	34	41	45	47	49	50	51	52	53
.59	36	43	47	49	50	52	52	53	54
.60	38	45	48	50	52	53	54	54	55
.61	40	46	50	52	53	54	55	56	56
.62	42	48	51	53	54	55	56	57	57
.63	44	49	52	54	56	57	57	58	58
.64	46	51	54	56	57	58	59	59	60
.65	48	53	55	57	58	59	60	60	61
.66	49	54	57	58	60	60	61	62	62
.67	51	56	58	60	61	62	62	63	63
.68	53	57	60	61	62	63	63	64	64
.69	55	59	61	62	63	64	65	65	65
.70	56	60	62	64	65	65	66	66	67
.71	58	62	64	65	66	66	67	67	68
.72	60	63	65	66	67	68	68	69	69
.73	61	64	66	67	68	69	69	70	70
.74	63	66	68	69	69	70	71	71	71
.75	64	67	69	70	71	71	72	72	72
.76	66	69	70	71	72	72	73	73	73
.77	67	70	71	72	73	74	74	74	75
.78	69	71	73	74	74	75	75	75	76
.79	71	73	74	75	76	76	76	77	77
.80	72	74	75	76	77	77	77	78	78

APPENDIX B

TABLE IV-11 (B)

Predictor $\underline{n}$ = 30

SHRUNKEN MULTIPLE-REGRESSION COEFFICIENTS

COMPUTED FROM WHERRY FORMULA

(See Chapter IV)

Discovered MULTR	Sample Size 100	125	150	175	200	225	250	275	300
.50	00	10	25	31	34	37	38	40	41
.51	0C	15	27	33	36	38	40	41	42
.52	00	19	29	34	37	40	41	43	43
.53	00	23	32	36	39	41	43	44	45
.54	00	26	34	38	41	43	44	45	46
.55	00	28	36	40	42	44	45	47	47
.56	12	31	37	41	44	46	47	48	49
.57	18	33	39	43	45	47	48	49	50
.58	22	35	41	45	47	48	50	50	51
.59	25	37	43	46	48	50	51	52	52
.60	29	39	45	48	50	51	52	53	54
.61	31	41	46	49	51	52	53	54	55
.62	34	43	48	51	52	54	55	56	56
.63	37	45	49	52	54	55	56	57	57
.64	39	47	51	54	55	56	57	58	59
.65	41	49	53	55	57	58	59	59	60
.66	44	51	54	56	58	59	60	61	61
.67	46	52	56	58	59	60	61	62	62
.68	48	54	57	59	61	62	62	63	63
.69	50	56	59	61	62	63	64	64	65
.70	52	57	60	62	63	64	65	65	66
.71	54	59	62	63	65	65	66	67	67
.72	56	60	63	65	66	67	67	68	68
.73	57	62	64	66	67	68	68	69	69
.74	59	64	66	67	68	69	70	70	71
.75	61	65	67	69	70	70	71	71	72
.76	63	67	69	70	71	72	72	73	73
.77	64	68	70	71	72	73	73	74	74
.78	66	70	71	73	73	74	74	75	75
.79	68	71	73	74	75	75	76	76	76
.80	69	72	74	75	76	76	77	77	77

APPENDIX B

TABLE IV-11 (C)

Predictor $\underline{n}$ = 35

SHRUNKEN MULTIPLE-REGRESSION COEFFICIENTS

COMPUTED FROM WHERRY FORMULA

(See Chapter IV)

Discovered MULTR	Sample Size 100	125	150	175	200	225	250	275	300
.50	00	00	14	25	29	33	36	37	39
.51	00	00	18	27	32	35	37	39	40
.52	00	00	22	29	34	37	39	40	42
.53	00	00	24	32	36	38	40	42	43
.54	00	11	27	34	37	40	42	43	44
.55	00	17	30	36	39	42	43	45	46
.56	00	21	32	38	41	43	45	46	47
.57	00	24	34	39	43	45	46	47	48
.58	00	27	36	41	44	46	48	49	50
.59	00	30	38	43	46	48	49	50	51
.60	10	33	40	45	47	49	51	52	52
.61	17	35	42	46	49	51	52	53	54
.62	22	38	44	48	50	52	53	54	55
.63	26	40	46	50	52	53	55	56	56
.64	29	42	48	51	53	55	56	57	58
.65	33	44	50	53	55	56	57	58	59
.66	36	46	51	54	56	58	59	59	60
.67	38	48	53	56	58	59	60	61	61
.68	41	50	55	57	59	60	61	62	63
.69	44	52	56	59	60	62	62	63	64
.70	46	54	58	60	62	63	64	64	65
.71	48	56	59	62	63	64	65	66	66
.72	51	57	61	63	64	66	66	67	67
.73	53	59	62	64	66	67	68	68	69
.74	55	61	64	66	67	68	69	69	70
.75	57	62	65	67	68	69	70	71	71
.76	59	64	67	69	70	71	71	72	72
.77	61	66	68	70	71	72	73	73	73
.78	63	67	70	71	72	73	74	74	75
.79	65	69	71	73	74	74	75	75	76
.80	67	71	73	74	75	76	76	77	77

APPENDIX B

TABLE IV-11 (D)

Predictor n = 40

SHRUNKEN MULTIPLE-REGRESSION COEFFICIENTS

COMPUTED FROM WHERRY FORMULA

(See Chapter IV)

Discovered MULTR	Sample Size 100	125	150	175	200	225	250	275	300
.50	00	00	00	16	25	29	33	35	37
.51	00	00	00	20	27	32	34	37	38
.52	00	00	05	23	29	33	36	38	40
.53	00	00	13	26	32	35	38	40	41
.54	00	00	18	28	34	37	40	41	43
.55	00	00	22	31	36	39	41	43	44
.56	00	00	25	33	38	41	43	44	46
.57	00	06	28	35	39	42	44	46	47
.58	00	14	30	37	41	44	46	47	48
.59	00	19	33	39	43	45	47	49	50
.60	00	24	35	41	45	47	49	50	51
.61	00	27	38	43	46	49	50	51	52
.62	00	30	40	45	48	50	52	53	54
.63	00	33	42	47	50	52	53	54	55
.64	10	36	44	48	51	53	54	56	56
.65	18	38	46	50	53	54	56	57	58
.66	23	41	48	52	54	56	57	58	59
.67	27	43	50	53	56	57	59	60	60
.68	31	45	51	55	57	59	60	61	62
.69	35	48	53	57	59	60	61	62	63
.70	38	50	55	58	60	62	63	63	64
.71	41	52	57	60	62	63	64	65	65
.72	44	54	58	61	63	64	65	66	67
.73	47	56	60	63	64	66	67	67	68
.74	49	58	62	64	66	67	68	69	69
.75	52	60	63	66	67	68	69	70	70
.76	54	61	65	67	69	70	70	71	72
.77	56	63	67	69	70	71	72	72	73
.78	59	65	68	70	71	72	73	74	74
.79	61	67	70	72	73	74	74	75	75
.80	63	68	71	73	74	75	76	76	76

APPENDIX B

TABLE IV-11 (E)

Predictor n = 45

SHRUNKEN MULTIPLE-REGRESSION COEFFICIENTS

COMPUTED FROM WHERRY FORMULA

(See Chapter IV)

Discovered MULTR	Sample Size 100	125	150	175	200	225	250	275	300
.50	00	00	00	00	18	25	29	32	34
.51	00	00	00	04	21	27	31	34	36
.52	00	00	00	13	24	29	33	36	38
.53	00	00	00	17	27	32	35	37	39
.54	00	00	00	21	29	34	37	39	41
.55	00	00	03	24	31	36	39	41	42
.56	00	00	13	27	34	38	40	42	44
.57	00	00	18	30	36	39	42	44	45
.58	00	00	22	32	38	41	44	45	47
.59	00	00	26	35	40	43	45	47	48
.60	00	00	29	37	42	44	47	48	50
.61	00	12	32	39	43	46	48	50	51
.62	00	18	34	41	45	48	50	51	52
.63	00	23	37	43	47	50	51	53	54
.64	00	27	39	45	49	51	53	54	55
.65	00	31	42	47	50	53	54	56	57
.66	00	34	44	49	52	54	56	57	58
.67	00	37	46	51	54	56	57	58	59
.68	12	40	48	52	55	57	59	60	61
.69	20	42	50	54	57	59	60	61	62
.70	25	45	52	56	58	60	61	62	63
.71	30	47	54	58	60	62	63	64	65
.72	34	49	56	59	61	63	64	65	66
.73	38	52	58	61	63	64	66	66	67
.74	41	54	59	62	64	66	67	68	68
.75	44	56	61	64	66	67	68	69	70
.76	48	58	63	66	67	69	70	70	71
.77	50	60	65	67	69	70	71	72	72
.78	53	62	66	69	70	71	72	73	73
.79	56	64	68	70	71	73	74	74	75
.80	58	66	70	72	73	74	75	75	76

APPENDIX B

TABLE IV-11 (F)

Predictor n = 50

SHRUNKEN MULTIPLE-REGRESSION COEFFICIENTS

COMPUTED FROM WHERRY FORMULA

(See Chapter IV)

Discovered MULTR	Sample Size 100	125	150	175	200	225	250	275	300
.50	00	00	00	00	00	19	25	29	32
.51	00	00	00	00	11	22	27	31	33
.52	00	00	00	00	16	25	30	33	35
.53	00	00	00	00	20	27	32	35	37
.54	00	00	00	08	23	30	34	37	39
.55	00	00	00	15	26	32	36	38	40
.56	00	00	00	19	29	34	38	40	42
.57	00	00	00	23	31	36	39	42	44
.58	00	00	04	26	34	38	41	43	45
.59	00	00	14	29	36	40	43	45	47
.60	00	00	19	32	38	42	45	47	48
.61	00	00	23	34	40	44	46	48	50
.62	00	00	27	37	42	46	48	50	51
.63	00	00	30	40	44	47	50	51	53
.64	00	10	33	41	46	49	51	53	54
.65	00	18	36	44	48	51	53	54	55
.66	00	23	39	46	50	52	54	56	57
.67	00	28	41	48	51	54	56	57	58
.68	00	31	44	50	53	55	57	59	60
.69	00	35	46	51	55	57	59	60	61
.70	00	38	48	53	56	59	60	61	62
.71	00	41	50	55	58	60	62	63	64
.72	16	44	52	57	60	61	63	64	65
.73	24	47	55	59	61	63	64	65	66
.74	29	49	56	60	63	65	66	67	68
.75	34	52	58	62	64	66	67	68	69
.76	38	54	60	64	66	68	69	70	70
.77	42	56	62	65	68	69	70	71	71
.78	46	59	64	67	69	70	71	72	73
.79	49	61	66	69	71	72	73	74	74
.80	52	63	68	70	72	73	74	75	75

<u>APPENDIX B</u>

TABLE IV-11 (G)

Predictor $\underline{n}$ = 55

SHRUNKEN MULTIPLE-REGRESSION COEFFICIENTS

COMPUTED FROM WHERRY FORMULA

(See Chapter IV)

Discovered MULTR	Sample Size 100	125	150	175	200	225	250	275	300
.50	00	00	00	00	00	08	19	25	28
.51	00	00	00	00	00	14	22	27	31
.52	00	00	00	00	00	18	25	30	33
.53	00	00	00	00	08	22	28	32	34
.54	00	00	00	00	15	25	30	34	36
.55	00	00	00	00	19	27	32	36	38
.56	00	00	00	00	23	30	35	38	40
.57	00	00	00	11	26	32	37	39	42
.58	00	00	00	17	29	35	39	41	43
.59	00	00	00	22	31	37	40	43	45
.60	00	00	00	25	34	39	42	45	46
.61	00	00	07	29	36	41	44	46	48
.62	00	00	16	32	39	43	46	48	50
.63	00	00	21	34	41	45	48	50	51
.64	00	00	25	37	43	47	49	51	53
.65	00	00	29	39	45	48	51	53	54
.66	00	00	32	42	47	50	53	54	56
.67	00	10	36	44	49	52	54	56	57
.68	00	18	38	46	51	54	56	57	58
.69	00	24	41	48	53	55	57	59	60
.70	00	29	44	50	54	57	59	60	61
.71	00	33	46	52	56	59	60	62	63
.72	00	37	49	54	58	60	62	63	64
.73	00	40	51	56	60	62	63	64	65
.74	00	43	53	58	61	63	65	66	67
.75	13	46	55	60	63	65	66	67	68
.76	22	49	57	62	65	66	68	69	69
.77	29	52	60	64	66	68	70	70	71
.78	34	54	62	65	68	69	71	71	72
.79	39	57	64	67	69	71	72	73	73
.80	44	59	66	69	71	72	73	74	75

APPENDIX C

COMPUTER PROGRAM

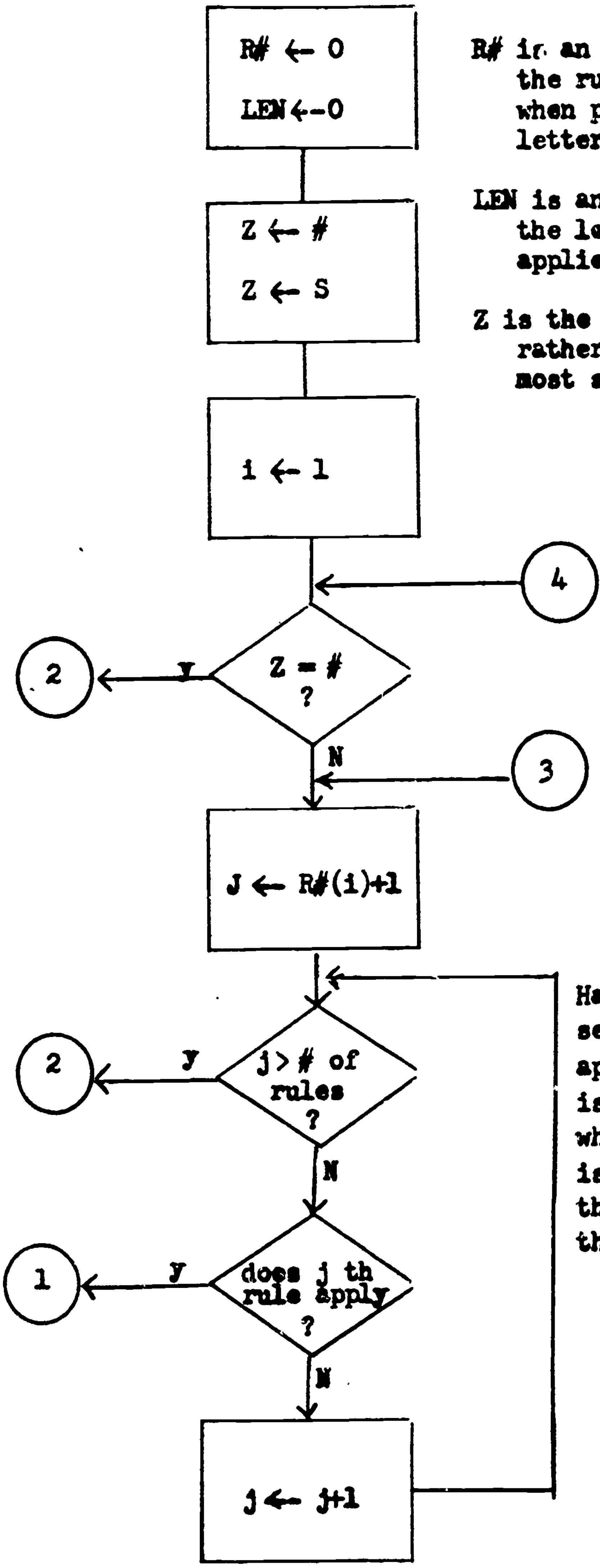
Written by Donald Marcotte

```
$IBFTC PHRAS
      SUBROUTINE PHRASE
      COMMON/INWD/INDPWD
      COMMON/IN/BROKUP,TEXT,LENGTH
      COMMON/PSUM/RELN,NEXT
      COMMON/MAPH/PHRMAT
      INTEGER RC (12),ITEXT(200),PHRMAT(300,8)
      COMMON/ACC/TOTAL,ID
      INTEGER THRZ,ASTRK,QUOTE
      COMMON/QTE/NQUOTE
      DATA PERIOA,EXCLAA,QUESTA/2H.*,3H.X*,3H.Q*/
      DATA COMMAA/1H,/
      INTEGER COMMAA,PERIOA,EXCLAA,QUESTA
      DATA ASTRK/1H*/
      DATA THRZS/3HZZZ/
      REAL HLFTXT(200)
      DOUBLE PRECISION TEXT(100),INDPWD(254)
      EQUIVALENCE(HLFTXT,TEXT,ITEXT)
      LOGICAL INTABL
      REAL BROKUP(80)
      INTEGER LENGTH(100),RELN(30),NEXT,TOTAL
      NQUOTE = 0
      LCSENT = 0
      QUOTE = 0
      IPE = 0
      LNEXT = 2*NEXT - 4
      DO 5 ISENT = 1,LNEXT
      IF(ISENT.LT.IPE) GO TO 5
      ICSENT = ISENT - ISENT/2
      IF(ICSENT.EQ.LCSENT) GO TO 5
      LCSENT = ICSENT
      IF(ITEXT(ISENT).EQ.COMMAA.OR.ITEXT(ISENT).EQ.PERIOA.OR.ITEXT(ISENT
     1).EQ.ASTRK) GO TO 5
      IF(INTABL(TEXT(ICSENT),INDPWD,508)) GO TO 118
      GO TO 5
  118 DO 4 IPB = 1,300
      IF(ITEXT(ISENT).EQ.PHRMAT(IPB,1)) GO TO 44
      GO TO 4
   44 ISPOW = ISENT + 1
      IF(ITEXT(ISPOW).EQ.PHRMAT(IPB,2)) GO TO 7
    4 CONTINUE
      GO TO 5
    7 IPE = ISPOW + 1
      KACC = 2
      RC(1) = PHRMAT(IPB,1)
      RC(2) = PHRMAT(IPB,2)
      IPC = IPB
      DO 13 LIPC = 3,8
```

```
      IF(ITEXT(IPE).EQ.PHRMAT(IPC,IIPC)) GO TO 21
      IF(PHRMAT(IPC,IIPC).EQ.THRZS) GO TO 23
      IF(KACC.EQ.2.OR.KACC.EQ.4.OR.KACC.EQ.6) GO TO 84
      GO TO 23
 21   KACC = KACC + 1
      RC(KACC) = PHRMAT(IPC,IIPC)
      IPE = IPE + 1
 13   CONTINUE
      GO TO 23
 84   IPB = IPB + 1
      IF(ITEXT(ISENT).EQ.PHRMAT(IPB,1)) GO TO 49
      IPE = ISENT + 2
      GO TO 5
 49   IF(ITEXT(ISPOW).EQ.PHRMAT(IPB,2)) GO TO 7
      IPE = ISENT + 2
      GO TO 5
 23   ICFAST = ISENT - 2
      LCFAST = IPE
      LWOP = KACC
      IF(ITEXT(ICFAST).EQ.ASTRK) GO TO 113
      GO TO 221
 113  IF(ITEXT(LCFAST.EQ.ASTRK) GO TO 114
      IF(ITEXT(LCFAST).EQ.COMMAA.OR.ITEXT(LCFAST).EQ.PERIOA.OR.ITEXT(LCF
     1AST).EQ.EXCLAA.OR.ITEXT(LCFAST.EQ.QUESTA) GO TO 114
      ICAC = LCFAST + 2
      IF(ITEXT(LCFAST).EQ.COMMAA.AND.ITEXT(ICAC).EQ.ASTRK) GO TO 114
      GO TO 221
 114  QUOTE = 1
      NQUOTE = NQUOTE + 1
 221  WRITE(7,223) ID,IPC,QUOTE,(RC(IRC),IRC=1,LWOP)
 223  FORMAT(5X,I5,5X,15,5X,I5,5X,8A6)
      TOTAL = TOTAL + 1
      WRITE(6,923) (RC(IRC),IRC = 1,LWOP)
 923  FORMAT(12HOPHRASE IS   ,12A6)
      DO 62 IRCA = 1,8
 62   RC(IRCA) = 0
      QUOTE = 0
  5   CONTINUE
      RETURN
      END
```

APPENDIX D

PL/I PROGRAM PARSE
Written by Gerald Fisher



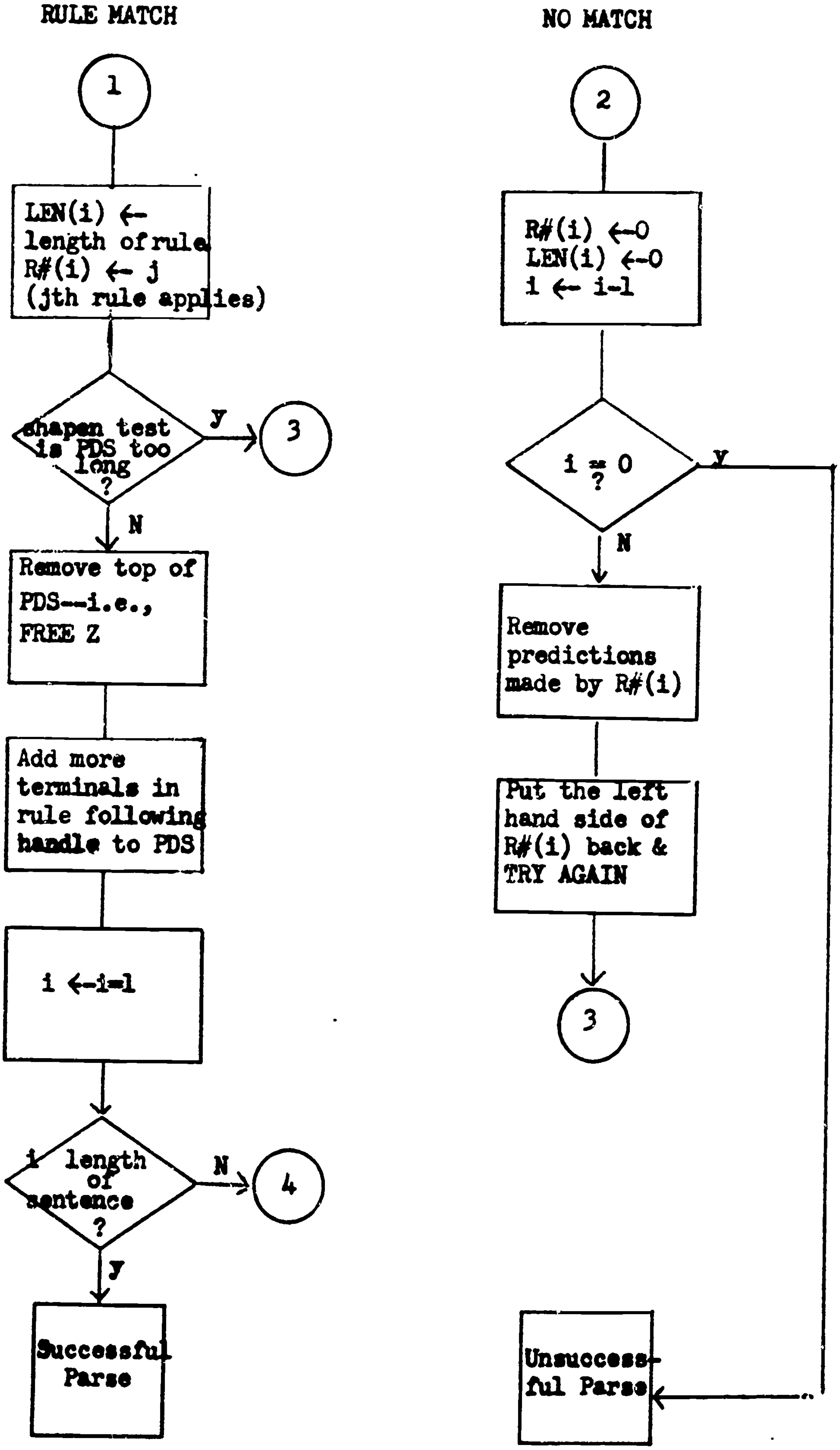
R# is an array. R#(i) is the rule number applied when parsing the i th letter.

LEN is an array. LEN(i) is the length of the rule applied to word i.

Z is the push-down store—rather Z gives the top-most symbol on the PDS.

Have we checked to see if any rule is applicable. —i.e. is there a rule whose left hand side is Z and handle is the i th word of the sentence?

APPENDIX D (Continued)



```
      /* A TEST OF THE PARSING PROCEDURE                                     */
 1    (CHECK(XX,RR)):TSTPAR:PROC OPTIONS(MAIN);
 2    DCL XX BIT(1);
 3    DCL R1(3,4) CHAR(2) INITIAL(
      'S','A','BB','#',           /* S->A BB       */
      'S','A','S','BB',           /* S-> A S BB    */
      'BB','B','#','#'),          /* BB-> B        */
      S1(4) CHAR(1) INITIAL('A','A','B','B'),  /* AABB */
      S2(3) CHAR(1) INITIAL('B','A','A'),      /* BAA */
      S3(7) CHAR(1) INITIAL('A','A','A','B','B','B','A'),  /* AAABBBA */
      R2(6,4) CHAR(2) INITIAL(
      'S','A','AA','#',           /* S->A AA       */
      'S','B','BB','#',           /* S-> B BB      */
      'S','A','S','AA',           /* S-> A S AA    */
      'S','B','S','BB',           /* S-> B S BB    */
      'AA','A','#','#',           /* AA-> A        */
      'BB','B','#','#'),          /* BB-> B        */
      S4(4) CHAR(1) INITIAL('A','B','B','A'),                 /* ABBA */
      S5(5) CHAR(1) INITIAL('B','A','B','B','B'),          /* BABBB */
      S6(6) CHAR(1) INITIAL('B','A','B','B','A','B');   /* BABBAB */
 4    DCL RR(10) FIXED;
 5    DCL R3(6,6) CHAR(5) INITIAL(
      'S','THE','ADJ','NOUN','V','ADJ',        /* S-> THE ADJ NOUN V ADJ */
      'S','THE','NOUN','V','ADJ','#',          /* S-> THE NOUN V ADJ */
      'NOUN','BOY','#','#','#','#',            /* NOUN-> BOY */
      'V','IS','#','#','#','#',                /* V-> IS */
      'ADJ','SMART','#','#','#','#',           /* ADJ-> SMART */
      'ADJ','HAPPY','#','#','#','#');          /* ADJ->HAPPY */
 6    DCL S7 (5) CHAR(5) INITIAL('THE','SMART','BOY','IS','HAPPY');
 7    PUT PAGE;
 8    CALL PARSE(S7,R3,RR,XX);
 9    PUT PAGE;
10    CALL PARSE(S1,R1,RR,XX);
11    CALL PARSE(S2,R1,RR,XX);
12    CALL PARSE(S3,R1,RR,XX);
13    CALL PARSE(S4,R2,RR,XX);
14    CALL PARSE(S5,R2,RR,XX);
15    CALL PARSE(S6,R2,RR,XX);
16    END TSTPAR;
```

```
   /* PARSE -- A PARSING PROCEDURE USING PREDICTIVE ANALYSIS.                  */
   /*THIS PROCEDURE WILL OBTAIN THE FIRST CORRECT PARSING OF AN INPUT          *
    *SENTENCE ACCORDING TO A SPECIFIED SET OF CONTEXT FREE REWRITING           *
    *RULES.  IT IS ASSUMED THAT THE GRAMMAR IS IN GREIBACH STANDARD FORM*
    *--I.E., THAT EACH IS OF THE FORM Y->AX1...XL (L>=0) WHERE Y,X1,...  *
    *XL ARE NON-TERMINAL AND A IS TERMINAL.                                    *
    *THE PARAMETERS ARE AS FOLLOWS:                                            *
    * SENTENCE IS THE INPUT SENTENCE TO BE PARSED.                             *
    * RULES      IS THE SET OF GRAMMAR RULES.                                  *
    * R#         IS THE PARSING TREE OUTPUT.                                   *
    * X          IS A SWITCH TO INDICATE WHETHER OR NOT THE PARSE WAS GOOD*
    *                                                                          *
    ***********************************************************************/
1   (CHECK(I,Z)):PARSE:PROC(SENTENCE,RULES,R#,X);
2   DCL SENTENCE(*) CHAR(*),(R#(*),LEN(DIM(SENTENCE,1))) FIXED,
    Z CHAR(*) CTL, /*NOTE THAT Z IS CONTROLLED;IT IS THE PDS */
    RULES(*,*) CHAR(*), /* THESE ARE THE RULES;EACH ROW IS ONE RULE */
    X BIT(1); /* IF THE PARSE IS SUCCESSFUL THEN X IS 1 AND 0 OTHERWISE */
3   N=DIM(SENTENCE,1); /*N GIVES US THE NUMBER OF WORDS IN SENTENCE */
4   M=DIM(RULES,1); /*M GIVES THE NUMBER OF RULES */
5   NN=DIM(RULES,2)-2;  /*NN GIVES THE MAX NUMBER OF NON-TERMINALS */
6   MM=LENGTH(RULES(1,1));   /* MM GIVES THE MAX WORD LENGTH */
7   R#=0;LEN=0;X='0'B; /* INITIALIZATION */
    /* FIRST WE SET UP THE PUSH DOWN STORE */
10  ALLOCATE Z CHAR(MM) INITIAL('#'); /*SYMBOL FOR BOTTOM OF STORE */
11  ALLOCATE Z CHAR(MM) INITIAL('S');
12  L=1; /* THIS IS A COUNTER FOR THE PDS */
13  DO I=1 TO N;  /* STEP THRU ONE WORD AT A TIME */
14  IF Z='#' THEN GO TO NO_MATCH;/*WE SHOULDN'T EMPTY THE PDS THIS SOON. */
    /* HERE WE SEE IF THERE IS A RULE WHOSE LEFT HAND SIDE MATCHES THE PDS
      AND WHOSE HANDLE IS THE CURRENT INPUT WORD.  BY KEEPING TRACK OF
      RULE NUMBERS WE MAKE SURE THAT WE DON'T TRY THE SAME RULE TWICE.    */
16  RULE_CHECK:DO J=R#(I)+1 TO M;
17  IF RULES(J,1)=Z&RULES(J,2)=SENTENCE(I) THEN GO TO MATCH;
19  END;
20  NO_MATCH:R#(I),LEN(I)=0; /* RESET IN CASE WE RETURN TO I */
21  I=I-1;  /* WE MUST BACK UP HERE */
22  IF I=0 THEN GO TO FREE_UP;  /*WE CAN DO NO MORE.  */
24  DO J=1 TO LEN(I);
25  FREE Z;  /*REMOVE THE PREDICTIONS MADE BY THIS RULE */
26  L=L-1;
27  END;
28  ALLOCATE Z CHAR(MM);Z=RULES(R#(I),1);  /*PUT THE GENERATRIX BACK */
30  L=L+1;
```

```
31   GO TO RULE_CHECK; /* MAYBE THERE IS ANOTHER WITH THE SAME HANDLE */
     /* AT THIS POINT WE HAVE A MATCH.  THUS, WE REMOVE THE TOP-MOST SYMBOL
       FROM THE PDS AND REPLACE IT WITH THE NON-TERMINAL PART OF THE RULE */
32   MATCH:R#(I)=J; /*REMEMBER THE RULE NUMBER. */
33   DO K=1 TO NN;
34   IF RULES(J,K+2)='#' THEN GO TO SET_LEN;
36   END;
37   SET_LEN:LEN(I)=K-1;
     /* SHAPER TEST--WILL THIS RULE REQUIRE MORE WORDS THAN REMAIN IN S?  */
38   IF L+K-2>N-I THEN GO TO RULE_CHECK;
40   FREE Z;   /* REMOVE THE TOP */
41   L=L-1;
     /* NOW ADD THE REMAINDER OF THE RULE TO THE PDS.            */


42    DO K=LEN(I) TO 1 BY -1;
43    ALLOCATE Z CHAR(MM);Z=RULES(J,K+2);
45    L=L+1;
46    END;
47    END; /* THIS ENDS THE MAIN ANALYSIS LOOP */
48    X='1'B; /*WE HAVE A SUCCESSFUL PARSE.  THE PDS MUST BE EMPTY */
49    FREE_UP:IF Z¬='#' THEN DO;FREE Z;GO TO FREE_UP;END;
54    FREE Z;
55    END PARSE;
```

ATTRIBUTE TABLE

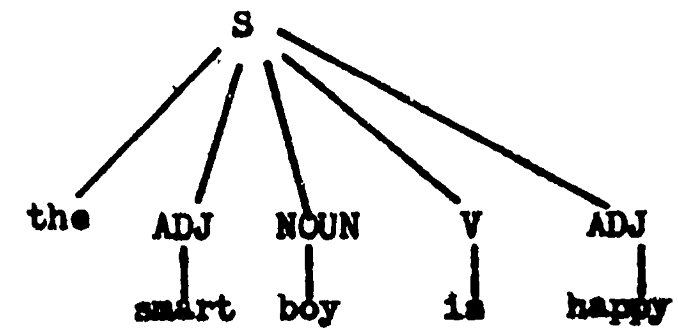
DCL NO.	IDENTIFIER	ATTRIBUTES
	DIM	GENERIC,BUILT-IN FUNCTION
49	FREE_UP	STATEMENT LABEL CONSTANT
	I	AUTOMATIC,BINARY,FIXED(15,0)
	J	AUTOMATIC,BINARY,FIXED(15,0)
	K	AUTOMATIC,BINARY,FIXED(15,0)
	L	AUTOMATIC,BINARY,FIXED(15,0)
2	LEN	(*),AUTOMATIC,DECIMAL,FIXED(5,0)
	LENGTH	GENERIC,BUILT-IN FUNCTION
	M	AUTOMATIC,BINARY,FIXED(15,0)
32	MATCH	STATEMENT LABEL CONSTANT
	MM	AUTOMATIC,BINARY,FIXED(15,0)
	N	AUTOMATIC,BINARY,FIXED(15,0)
	NN	AUTOMATIC,BINARY,FIXED(15,0)
20	NO_MATCH	STATEMENT LABEL CONSTANT
1	PARSE	ENTRY,DECIMAL,FLOAT(SINGLE)
2	R#	(*),PARAMETER,DECIMAL,FIXED(5,0)
16	RULE_CHECK	STATEMENT LABEL CONSTANT
2	RULES	(*,*),PARAMETER,ALIGNED,STRING,CHARACTER
2	SENTENCE	(*),PARAMETER,ALIGNED,STRING,CHARACTER
37	SET_LEN	STATEMENT LABEL CONSTANT
2	X	PARAMETER,STRING,BIT
2	Z	CONTROLLED,STRING,CHARACTER

NO ERRORS OR WARNINGS DETECTED.

COMPILE TIME .12 MINS

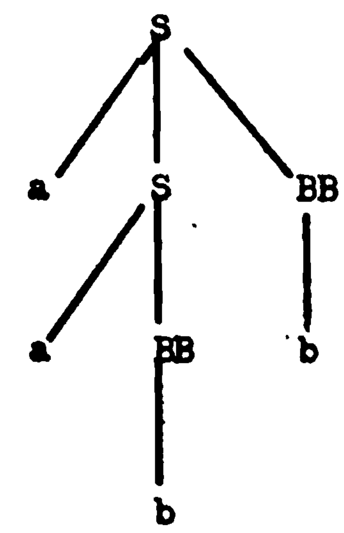
Sentence = the smart boy is happy.

```
I =          1;  [S]  [the]
Z = 'ADJ  ';        So it predicts
Z = 'V    ';        Adj noun v adj
Z = 'noun ';        (rule 1)
Z = 'ADJ  ';
```




```
I =          2;  Apply the rule   Adj ---> smart     (rule 5)
I =          3;  Apply the rule   Noun ---> boy      (rule 3)
I =          4;  Apply the rule   V ---> is          (rule 4)
I =          5;  Apply the rule   Adj ---> happy     (rule 6)
I =          6;  End of sentence.

RR(1)=  1    RR(2)=  5    RR(3)=  3    RR(4)=  4    RR(5)=  6
RR(6)=  0    RR(7)=  0    RR(8)=  0    RR(9)=  0    RR(10)=  0;

XX='1'B;

             RR(5)=  6
             RR(10)=  0;
```

Sentence = aabb




```
I=      1;

Z='BB';

I=      2;

I=      1;

Z='S ';

Z='BB';

Z='S ';

I=      2;

Z='BB';

I=      3;

I=      4;

I=      5;

RR(1)=  2     RR(2)=  1     RR(3)=  3     RR(4)=  3     RR(5)=  0
RR(6)=  0     RR(7)=  0     RR(8)=  0     RR(9)=  0     RR(10)=  0;
RR(5)=  0
RR(10)=  0;  PARSED
XX='1'B;
              Sentence = baa
I=      1;

I=      0;

RR(1)=  0     RR(2)=  0     RR(3)=  0     RR(4)=  0     RR(5)=  0
RR(6)=  0     RR(7)=  0     RR(8)=  0     RR(9)=  0     RR(10)=  0;
              RR(5)=  0
XX='0'B;      RR(10)=  0;
      NOT PARSED
```

Sentence = aaabbba

I= 1; |S| |a|

Z='BB'; |BB| |a| Try rule S → a BB

I= 2; Doesn't work--no rule of the form BB → $aY_1 \to Y_n$

I= 1; Back up one letter. Remove BB.

Z='S '; Put S back on PDS |S| |a|

Z='BB'; Try the rule S → a S BB

Z='S '; Now PDS contains |S BB|

I= 2; Second letter again. The pair is |S BB| |a|

Z='BB'; Try the rule S → a BB Hence PDS has |BB BB| on it.

I= 3; Use the rule BB → b Doesn't work

I= 2; Use the rule S → a S BB

Z='S '; Now PDS contains |S BB BB|

Z='BB';

Z='S ';

I= 3; Try third letter again

Z='BB'; Use the rule S → a BB

I= 4; Use the rule BB → b

I= 5; Use the rule BB → b

I= 6; Use the rule BB → b Now the PDS is empty

I= 7; a is input PDS is empty

I= 6; Back up.

Z='BB'; Put back on BB

I= 5; No rule Back up

Z='BB'; Put back on BB

I= 4; No rule back up

Z='BB'; Put back BB

I= 3; No rule try again

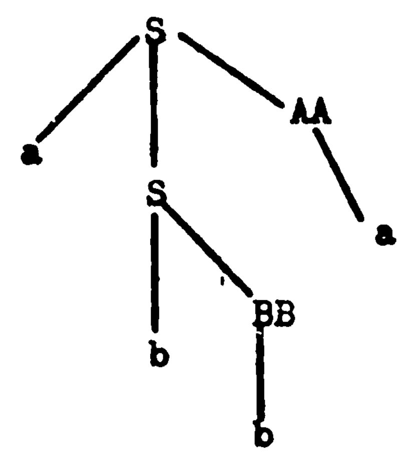
Z='S '; Put back S.

Z='BB'; Try the rule S → a S BB

Z='S ';

```
I=        4;   No prediction for
I=        3;   Back up.
Z='S ';        Put back S.
I=        2;   No rule left back up
Z='S ';        Put back S
I=        1;   Back up
Z='S ';        Put back S
I=        0;   That's all!
RR(1)=    0      RR(2)=   0      RR(3)=   0      RR(4)=   0      RR(5)=   0
RR(6)=    0      RR(7)=   0      RR(8)=   0      RR(9)=   0      RR(10)=   0;
                 RR(5)=   0
XX='0'B;         RR(10)=   0;
```

Sentence = ABBA

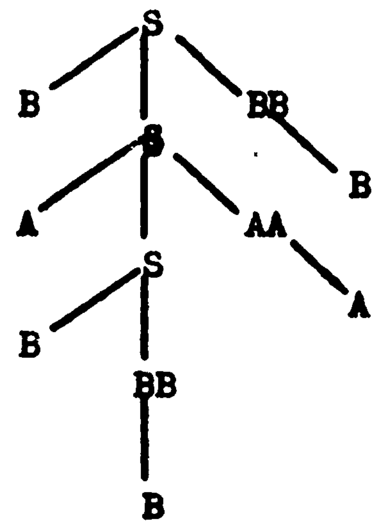



```
I=        1;
Z='AA';
I=        2;
I=        1;
Z='S ';
Z='AA';
Z='S ';
I=        2;
Z='BB';
I=        3;
I=        4;
I=        5;
RR(1)=    3      RR(2)=   2      RR(3)=   6      RR(4)=   5      RR(5)=   0
RR(6)=    0      RR(7)=   0      RR(8)=   0      RR(9)=   0      RR(10)=   0;
                 RR(5)=   0
XX='1'B;         RR(10)=   0;
```

```
I=        1;                   Sentence = BABBB

Z='BB';                        Failed because of the shaper test

I=        2;

I=        1;

Z='S ';

Z='BB';

Z='S ';

I=        2;

Z='AA';

I=        3;

I=        2;

Z='S ';

Z='AA';

Z='S ';

I=        3;

I=        2;

Z='S ';

I=        1;

Z='S ';

I=        0;

RR(1)=   0    RR(2)=   0    RR(3)=   0    RR(4)=   0    RR(5)=   0
RR(6)=   0    RR(7)=   0    RR(8)=   0    RR(9)=   0    RR(10)=   0;
              RR(5)=   0
XX='0'B;      RR(10)=   0;
```

```
I=        1;
Z='BB';
I=        2;
I=        1;
Z='S ';
Z='BB';
Z='S ';
I=        2;
Z='AA';
I=        3;
I=        2;
Z='S ';
Z='AA';
Z='S ';
I=        3;
Z='BB';
I=        4;
I=        5;
I=        6;
I=        7;
RR(1)=  4     RR(2)=  3     RR(3)=  2     RR(4)=  6     RR(5)=  5
RR(6)=  6     RR(7)=  0     RR(8)=  0     RR(9)=  0     RR(10)=  0;
              RR(5)=  5
XX='1'B;      RR(10)=  0;
```

Sentence = BABBAB



References

Baker, Frank B. Experimental design considerations associated with large scale research projects. Paper presented at the Seventh Annual Phi Delta Kappa Symposium on Educational Research, Madison, Wisconsin, August 11, 1965.

Becker, J., and Hays, R. N. Information Storage and Retrieval: Tools, Elements, Theories. New York: Wiley, 1963.

Bobrow, D. G. "Syntactic Theories in Computer Implementations," in Harold Borko (Ed.), Automated Language Processing. New York: Wiley, 1967. Pp. 215-251.

Booth, T. L. Sequential Machines Automata Theory. New York: Wiley, 1967.

Borko, Harold (Editor). Automated Language Processing. New York: Wiley, 1967.

Borko, Harold (Editor), Computer applications in the behavioral sciences. Englewood Cliffs, N. J. : Prentice-Hall, 1962, 633p.

Bowles, Edward. Computers in Humanistic Research. Englewood Cliffs, N. J.: Prentice-Hall, 1967. Pp.264.

Carne, E. B. Artificial Intelligence Techniques. Washington, D.C.: Spartan, 1965.

Carroll, John B. Linguistics and the psychology of language. Review of Educational Research, 1964, 34, 119-126.

Chomsky, Noam. Syntactic structures. "s-Gravenhenge, Netherlands: Mouton and Company, 1957. 116 p.

Christal, R. E. "Jan: A Technique for Analyzing Group Judgment." Technical Documentary Report PRL-TDR-63-3. February 1963. Project 7734, Air Force Systems Command, Lackland Air Force Base, Texas.

Daigon, Arthur. Computer Grading of English Composition. The English Journal, January, 1966, 46-52.

Dale, Edgar, and Chall, Jean. A formula for predicting readability. Educational Research Bulletin, 1948.

Diederich, Paul B., French, John W., and Carlton, Sydell T. Factors in judgments of writing ability. ETS Research Bulletin, 1961, No. 15. Pp.60.

Doyle, L. B. Seven ways to inhibit creative research. Datamation, February, 1965.

Fowler, H. W. A Dictionary of Modern English Usage. Second edition revised by Sir Ernest Gowers. London: Oxford University Press, 1965. 725 p.

Friedman, N. and McLaughlin, C. Logic, rhetoric and style. Boston: Little, Brown, 1963. Pp.230.

Garber, H. and Shostak, R. "Judge Viewpoints in Grading Essays." Paper presented at the Annual Meeting of the American Psychological Association, Washington, D. C., September 2, 1967.

Garvin, Paul L. (Ed.) Natural language and the computer. New York: McGraw-Hill, 1963. 398p.

Gates, Arthur H. Spelling difficulties in 3876 words. New York: Bureau of Publications, Teachers College, Columbia University, 1937.

Gleason, H. A., Jr. What grammar? Harvard Educational Review, 1964, 34, 267-281.

Gleason, H. A., Jr. Linguistics and English Grammar. New York: Holt, Rinehart and Winston, 1965. Pp.519.

Green, Bert F., Jr. Digital computers in research. New York: McGraw-Hill, 1963. 333p.

Guth, Hans P. English Today and Tomorrow. Englewood Cliffs, N. J.: Prentice-Hall, 1964. Chapter 4, p. 165-217.

Hays, David G. Introduction to Computational Linguistics. New York: American Elsevier, 1967. pp.231.

Hiller, Jack H., Page, E. B. and Marcotte, D. R. A Computer Search for Traits of Opinionation, Vagueness, and Specificity-Distinction in Student Essays. Paper read at the Annual Meeting of the American Psychological Association, Washington, D. C., September 2, 1967.

Hodges, John C. College Handbook, 3rd. ed. New York: Harcourt, Brace, 1951. 494 p.

Hunt, Kellogg, W. "Recent Measures in Language Development." Paper read to National Conference on Research in English. AERA Annual Meetings, Chicago, Illinois, February 19, 1966.

Jackson, D. N. and Messick, S., "Individual differences in social perceptions," Brit. J. Social and Clinical Psychology, 1963, 2, Pp. 1-10.

Kelley, T. L. Fundamental Statistics. Cambridge: Harvard University Press, 1947.

Keyser, S. J., and Petrick, S. R. Syntactic Analysis, 1966. (In press in a forthcoming book.)

Klein, S. and Simmons, R. F. A computational approach to grammatical coding of English words. Journal of Association of Computing Machinery. 1963, 10, 334-347.

Kuno, Susumu. Some characteristics of the Multiple-Path Syntactic Analyzer. Language Data Processing, Cambridge: Harvard Computation Laboratory, 1964. C6, 1-8.

Kuno, S. and Oettinger, A. Mathematical Linguistics and Automatic Translation. Report No. NSF-14. Cambridge, Mass.: Harvard Computation Laboratory, 1965.

Korfhage, Robert R. Logic and Algorithms: With Applications to the Computer and Information Sciences. New York: Wiley and Sons, 1966. Pp. 194.

LaBrant, Lou L. "Analysis of Cliches and Abstractions." English Journal 38: 275-78; May 1949.

Lawson, H. W. PL/I List Processing. Communications of the ACM, 1967, 10(6), 358-367.

Levien, R. E. and Maron, M. E. "A Computer System for Inference Execution and Data Retrieval." Communications of the ACM, 1967, 10(11), 715-721.

Loban, Walter. The language of elementary school children. Champaign, Ill.: National Council of the Teachers of English, 1963.

Lorge, Irving. The Lorge Formula for estimating difficulty of reading materials. New York: Bureau of Publications, Teachers College, Columbia University, 1959, 20 p.

Marcotte, Donald. A Computer Analysis of Cliche Behavior. M.A. Thesis, University of Connectiqut, 1966.

Marcotte, D. R., Page, E. B., and Daigon, A. A Computer Analysis of Cliche Behavior in Student Prose. Paper read at the Annual Meeting of the American Educational Research Association, New York, February 18, 1967.

Marksheffel, Ned D. Composition, handwriting, and spelling. Review of Educational Research, 1964, 34, 177-186.

McColly, William and Remstad, Robert. Comparative effectiveness of composition skills learning activities in the secondary school. USOE Cooperative Research Project 1528. Madison, Wisc.: Univ. of Wisconsin, 1963. 94 p.

McNemar, Quinn. Psychological Statistics, 3rd ed. New York: Wiley, 1962.

Merwin, Jack C. and Gardner, Eric F. Development and application of tests of educational achievement. Review of Educational Research, 1962, 32, 40-50.

Miller, George A. Some psychological studies of grammar. American Psychologist, 1962, 17, 748-762.

Minsky, M. L. Computation: Finite and Infinite Machines. Englewood, N.J.: Prentice-Hall, 1967.

Myers, Albert B., McConville, Carolyn B., and Coffman, William E. Reliability of reasoning of the College Board English Composition Test, December 1963. ETS Research Bulletin, 1964, No. 42. 22p.

Oettinger, A. G. Automatic Language Translation. Cambridge, Mass.: Harvard University Press, 1960.

Ogilvie, D. M., Dunphy, D. C., Smith, C., Stone, P. J., with Shneidman, E., and Farberow, N. Some characteristics of genuine versus simulated suicide notes as analyzed by a computer system called the General Inquirer. Cambridge, Mass.: Harvard University Laboratory of Social Relations, 1962. Ditto.

Page, Ellis B. The Imminence of Grading Essays by Computer. Phi Delta Kappan, January, 1966, 238-243.

Page, Ellis B. Grading Essays by Computer: Progress Report. Proceedings of the 1966 Invitational Conference on Testing Problems. Princeton, N.J.: Educational Testing Service, 1967a. Pp. 87-100.

Page, Ellis B. "Statistical and Linguistic Strategies in the Computer Grading of Essays." Proceedings of the Second International Conference on Computational Linguistics. Grenoble, France. August 24, 1967. No. 34.

Partridge, Eric. A Dictionary of Cliches, with an Introductory Essay. Reprinted (with some additions). London: Routledge and Kegan Paul Ltd., 1962. 260 p.

Paulus, Dieter and Page, Ellis B. Nonlinearity in Grading Essays. Paper read at the Annual Meeting of the American Educational Research Association, New York, February 16, 1967.

Paulus, Dieter. Feedback in Project Essay Grade. Paper read at the Annual Meeting of the American Psychological Association, Washington, D.C., September 2, 1967.

Pence, R. W. A Grammar of Present-Day English. New York: Macmillan, 1947.

Postal, Paul M. Underlying and superficial linguistic structure. Harvard Educational Review, 1964, 34, 246-266.

Puckett, Arlette L. STUFF: String Utility Routines for FORTRAN IV. SHARE Program Library, 1966, S.D.A. 3454.

Quillian, M. Ross. Semantic Memory. Cambridge, Mass.: Bolt Beranek and Newman, 1966.

Sakoda, James M. DYSTAL manual: Dynamic Storage Allocation Language in FORTRAN. Providence: Brown University Sociology Computer Laboratory, 1964. 195 p. + app. Mimeograph.

Scott, T. M. Composition for College Students. New York: Macmillan, 1960.

Simon, Herbert A. Simulation of human thinking. In Martin Greenberger (Editor), Management and the computer of the future. Cambridge, Mass.: Massachusetts Institute of Technology Press, 1964. Pp. 94-131.

Stolz, Walter S., Tannenbaum, Percy H., and Carstensen, Frederick V. A stochastic approach to the grammatical coding of English. University of Wisconsin Communications Center, 1965? Ditto, 23 Pp.

Stone, Philip. Paper read to the Conference on Uses of Educational Data Banks, Boston, Mass., Dec. 4, 1964. Unpublished.

Stone, Philip J., and Hunt, Earl B. A computer approach to content analysis: Studies using the General Inquirer system. Proceedings of the Spring Joint Computer Conference, 1963, 241-256.

Stone, Philip J., Dunphey, Dexter C., Smith, Marshall S., and Ogilvie, Daniel M. The General Inquirer: A Computer Approach to Content Analysis. Cambridge: M.I.T. Press, 1966. Pp. 651.

Strunk, W., Jr., and White, E. B. The Elements of Style. New York: Macmillan, 1959.

Summey, George Jr. American Punctuation. New York: Ronald Press Co., 1949.

Tatsuoka, M. and Tiedeman, D. V. "Statistics as an Aspect of Scientific Method in Research on Teaching." In Gage, N. L. (Ed.), Handbook of Research on Teaching. Chicago: Rand McNally, 1963. Ch. 4.

Warriner, John E. Handbook of English. New York: Harcourt, Brace, 1951. 594 p.

References (Continued)

Warriner, John E. and Griffith, Francis. English Grammar and Composition. New York: Harcourt, Brace and Co., 1957.

Weizenbaum, J. Symmetric List Processor. Communications of the Association for Computing Machinery, 1963, 6(9), 524-536.

Woods, William A. Semantics for a Question-Answering System. Paper read at the Annual Meeting of the Association for Machine Translation and Computational Linguistics. Atlantic City, N.J. April 21, 1967.

Yngve, V. An introduction to COMIT programming. Cambridge, Mass.: Institute of Technology Press, 1962a. 62 p.

Yngve, V. COMIT programmers' reference manual. Cambridge, Mass.: Massachusetts Institute of Technology Press, 1962b. Pp. 61.