A COMPUTER GENERATED CORPUS AND LEXICAL ANALYSIS OF ENGLISH LANGUAGE INSTRUCTIONAL MATERIALS PRESCRIBED FOR USE IN BRITISH COLUMBIA JUNIOR SECONDARY GRADES

by

PETER EDWARDS

B.A., University of Western Australia, 1963
B.Ed., University of Western Australia, 1967
M.A., University of British Columbia, 1972

A THESIS SUBMITTED IN PARTIAL FULFILMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF EDUCATION

in the Department

of

READING EDUCATION
FACULTY OF EDUCATION

We accept this thesis as conforming to the
required standard

Adviser

External
Examiner

THE UNIVERSITY OF BRITISH COLUMBIA

October, 1974

Department of **Faculty of Education.**

The University of British Columbia
Vancouver 8, Canada

Date October, 1974.

ABSTRACT

The major purpose of the study was to capture a representative sample of natural language from the textbooks prescribed for use in the junior secondary curriculum for British Columbia schools, organize the sample for computer processing through the development of needed programs, develop a lexical analysis and describe the word and sentence characteristics of the samples organized by grades, subjects across grades, subjects within grades and textbook corpora. A number of hypotheses related to the distribution of frequently occurring words and a sub-set of representative sentence lengths across the corpora were then tested and a model was developed to aid in selecting lexically significant vocabulary from word lists based on samples from subject area textbooks.

A stratified sampling model, applied to thirty-seven textbooks from seven subject areas, produced a Corpus of approximately a quarter million running words of natural language text based on 469 samples of approximately 500 words each. The results of the lexical analysis indicated that Grade 9 makes significantly greater reading demands in terms of volume of material (tokens) and vocabulary (word-types) than either Grades 8 or 10. Considerable diversity was exhibited in type and token distribution by grades, subjects, and textbooks but no apparent pattern emerged. However, use of Yule's K characteristic to determine the repeat rate frequency of word-types across the various corpora, revealed great variation in redundancy of word-types with the most striking differences exhibited in the samples

from English textbooks and to some extent those from Home Economics and Commerce. Similar results were obtained in applying Yule's K as a measure of the repeat rate frequency for sentence lengths. Samples from English textbooks, again, exhibited exceptional variability in sentence length variety. These results were further substantiated by the analysis of other measures of variability based on computation of standard deviations, coefficients of variation, Pearson's skew factor and, to a lesser degree, the average number of sentences per 500 word sample. In all instances, organization of the samples by gross grade groupings tended to mask the real inherent variability of the samples organized by subjects and textbooks.

Chi-square analyses of word and sentence distribution further substantiated the inherent variability revealed by the lexical analysis. Little uniformity was exhibited in the distribution of the most frequently occurring words in English and a representative sub-set of sentence lengths with the samples organized by grade levels, subjects across grades and subjects within grades. Grouping by gross grade level again masked subject variations. The style and content characteristics of the print materials prescribed for use in the separate subject areas are therefore significantly instrumental in affecting the frequency of occurrence of even the most common words in English and a representative sub-set of sentence lengths.

Further analysis of the word lists produced in the study substantiated the utility of developing an elimination technique, based on omission of the most frequently occurring words and the relatively rare words, to identify the significant vocabulary from word lists based on samples from texts in subject areas.

The major conclusion of the study suggests that the print materials prescribed for use in junior secondary grades exhibit marked variability when examined on even the most straightforward of linguistic characteristics such as word and sentence frequency. It is suggested that this variability would be even more pronounced if analyses were developed based on other syntactic and semantic variables. The expertise of the subject area specialist and the reading specialist should be combined in developing instruction to maximize learning from print materials. Such instruction would best be based on materials organized by subjects across grades and by separate subjects within grades rather than on materials organized by gross grade groupings.

ACKNOWLEDGEMENTS

I wish to express my sincere gratitude to the following people for their role in the completion of this dissertation.

Dr.E.G.Summers, my adviser and thesis supervisor, for his help in initially defining the problem, and for his guidance, counsel, patience, and encouragement during the preparation of the dissertation.

The members of my thesis committee, Dr.R.Bentley, Dr.J.Catterson, Dr.Br.L.Courtney, and Dr.D.Pratt, for their advice, support, and many fruitful suggestions made throughout the course of the study.

Dr.J.Bormuth of the University of Chicago, for the time he spent examining the dissertation and the helpful suggestions he made.

Mr.J.Coulthard, Mr.A.Miller, Inger Nissen and Irene Amiraslany of the Computing Centre, for their help and advice during the computer programming and key-punching of the material used in the study. I would especially like to thank Allan Miller for his enthusiasm and expertise in developing the numerous computer techniques and programs required during the study.

Finally, I would like to thank Dr.E.N.Ellis of the Vancouver School Board, for his help and kindness in obtaining the textbooks used in the dissertation.

TABLE OF CONTENTS

TABLE OF CONTENTS

TABLE OF CONTENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF TABLES

LIST OF TABLES

LIST OF TABLES

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDIXES

CHAPTER I

THE PROBLEM

BACKGROUND OF THE PROBLEM

Research into the various constituents of the reading act suggests that development of reading proficiency is a cumulative, life-long process requiring continued learning and refinement. Few would argue with the statement that, "The ability to read well constitutes one of the most valuable skills a person can acquire. Our world is a reading world" (Bond and Tinker, 1967). The use of written language gives man a permanent, external memory in striking contrast to the more ephemeral spoken word. In fact, reading instruction could be defined as the socially planned, guided or aided establishment of competency in dealing with the external print memory system of man.

Unlike certain characteristics that are genetic in origin, reading proficiency is generally an acquired skill which must be relearned by each generation. Thus, developing skill in understanding printed language has been a basic objective in man's educative processes since early recorded history (Dodds, 1967). Einstein suggested that the ability to read was mankind's most amazing feat. In an address presented to the 1972 convention of the International Reading Associaticn, Jerome Bruner (1972) noted that the capacity to read and extract meaning from the printed page may represent the ultimate stage in the evolution of homo sapiens. Comprehending the printed page is the result of complex interaction between physiological and psychological processes still but vaguely understood.

With the increased demand for literacy as technology and information have increased, reading instruction, cnce an exclusive concern of the elementary school, has become an important area of study in secondary schools. Evidence of its importance is readily available from a variety of sources.

An information base of research on reading has existed for seventy-five years. Such research has explored topics related to: sequencing and developing reading instruction at all levels, the process of reading, the products or skills of reading, language development as it relates to reading, the pedagogy of reading, and the special problems of the disabled reader (Robinson et al, 1967; Summers et al, 1968, 1967, 1968). Roughly 40 percent of the reported research relates to reading beyond the elementary level.

A significant trend in secondary reading instruction in recent years has been the attention given to more systematic development of reading abilities including the organization of special reading programs, increased emphasis on reading as it relates to subject classes, and provision of special services for students with serious reading problems (Davis, 1952; Robinson et al, 1960; Summers, 1963, 1964, 1966, 1967; Artley, 1968, 1970; Farr et al, 1970; Hill and Bartin, 1971).

In addition, the volume of interest is evidenced by the over fifty sources cited in a recent bibliographical guide, indexing information sources for secondary reading (Summers et al, 1973). In recent years, there has also been a massive outpouring of offerings from over 200 North American publishers developing instructional materials specifically designed for student use in reading instruction programs ranging from preschool through college and adult levels.

The trend towards more interest in secondary reading has some old and some new facets. A facet that reflects a reasonably old interest is that of focussing on the linguistic analysis of print material. The most notable and extensive linguistic analyses which have generated word lists and results applicable to instructional settings and school materials have been those reported by Thorndike and Lorge (1944), Rinsland (1945), Carroll et al (1971), and the study by Harris and Jacobson (1972). Although based on a sampling of more general adult materials, the project reported by Kucera and Francis (1967) represented the first study of a massive million word, computer-based corpus which generated results based on word and sentence lengths. The techniques employed are relevant to any analysis of school based instructional materials.

In recent research, computer technology has provided an important tool for analyzing transformed natural language text, organizing corpora, developing word frequency counts, and more significantly, enabling the analysis and comparison of masses of data across numerous sub-components within a corpus of materials. Computerized text data bases facilitate the use of varied statistical analyses and make it possible to study the linguistic characteristics of sizable bodies of written language. The advantages of computer technology in linguistic research are aptly illustrated by Kucera (1969).

> Since any useful analysis of language usage has to be based on a large body of textual material, even elementary information could be obtained, before the advent of computers, only with enormous labor. Let us imagine that one wished to determine some very basic lexical properties of a textual corpus containing a million running words. If this were to be done by hand (or, more accurately, by the human brain), the task would require an inordinate amount of time; each of the one million words would have to be inspected individually, and each new word recorded after first checking to make sure it had not already been noted. If the analysis were also to preserve information about the frequency of occurrence of individual words, or perhaps references to the pages or lines of the text where their occurrences were to be found, the assignment would become more formidable still..... Linguists and lexicographers alike have found in the computer a new and useful tool that has not only made the analysis of language less time-consuming but has also opened new insights into important problems in language usage.

New avenues to research have also been opened by an upsurge in interest in the linguistics of written language and the readability of instructional materials. There is great interest in frequency counts and item co-occurrence, positional criteria based on the placement of items within a text, syntactic criteria based on structural relationships between items, and semantic criteria depending on the particular area of discourse and on the larger context within which a given text is placed.

As Robinson (1971) pointed out:

> "We must study our language before we generate approaches to reading instruction....we need to learn more about the patterns of specific letters in words, sentence patterns, and the overall organizational patterns of our language."

In a recent article, Jenkinson (1970) analyzed current information gaps related to research on reading comprehension, and outlined vital areas needing further study including "problems inherent within the materials." Jenkinson concluded her discussion by stressing the need for "further linguistic analysis of the language of textbooks..."

The work of Smith (1964) was a major attempt to subjectively identify patterns of writing in different subject areas and to relate necessary reading skills to these patterns. The analysis included reading and study skills which were common to texts in Literature, Social Studies, Science and Mathematics in Grades 7 through 12.

If the instructional materials selected for use in a school program reflect society's communication with students through the language of print, then it becomes increasingly important to ask, "What are the linguistic characteristics of the print sources prescribed for use in Canadian secondary schools?" Answers to this query may well form the basis for instruction that is better adjusted to the real reading ability and needs of secondary students.

This question provides the basis for the present study which was undertaken as a contribution to research reflecting old and new trends in two main areas: 1) the analysis of certain linguistic characteristics of a sample of natural language text which forms the basis of an existing secondary school curriculum, and 2) the application of computer techniques to the analysis of natural language text.

STATEMENT OF THE PROBLEM AND PURPOSE OF THE STUDY

Numerous studies reported in the literature of linguistics and education illustrate attempts to describe and compare the language content of a body of print materials. The basic problems in this study were to describe and compare certain linguistic features across and within the grades and subject areas comprising a sample of printed instructional materials prescribed for use in the various subject areas of Grades 8, 9 and 10 in British Columbia through the development of a model utilizing computer technology.

The specific purposes were to generate a natural language corpus and to make various linguistic analyses (involving word frequency and sentence lengths) and comparisons of the total corpus and its sub-components by applying the power of computer storage and programming techniques.

The ideal study in describing the linguistic characteristics of print encountered by Grade 8, 9 and 10 students would draw samples from all possible print sources the student comes in contact with, including regular textbooks, supplementary sources, reference materials, and perhaps even samples of student written and spoken language. However, this study concentrates on a single print component and analyzes a limited number of explicitly defined, carefully selected, readily available language samples from the text materials that students are most likely to encounter during their junior secondary school years in subject classes. The print materials used for analysis in this study consisted of approximately a quarter million running words of natural language systematically sampled from texts prescribed for use in subjects in Grades 8, 9 and 10 in British Columbia schools.

TASKS, QUESTIONS AND HYPOTHESES

A complete description of instructional materials would be built around clearly isolated linguistic variables identified in a fully developed theory of language comprehension. However, such a theory has yet to be developed. Innumerable variables relate to comprehensibility of printed materials. In a recent seminal study, Bormuth (1969) analyzed factors that relate to the comprehension of print materials and identified 169 linguistic variables that correlate with comprehensibility and readability.

The state-of-the-art is such that it is not yet possible to explicate a complete theoretical account of the comprehension process, determine the linguistic variables in print materials which correlate most closely with comprehension difficulty, and develop pedagogical procedures and instructional materials that are consistently predictive in generating high level student comprehension of print materials.

The development of a fully explicated, scientific theory of language comprehension will no doubt emerge gradually. Until a fully developed theory can be used to generate studies, research based on straightforward manipulable variables that evidence indicates are related to the comprehension process, can provide important insights into the factors in instructional materials which may contribute to their diverse comprehension demands. Such descriptive and comparative research can produce results which may influence pedagogy and increase effectiveness in teaching students to acquire knowledge from written instructional materials.

In this study, the focus is on describing and comparing word and sentence characteristics of instructional materials prescribed for use in seven secondary subject areas in British Columbia schools.

The descriptive and comparative analyses are framed within a stratification model allowing data to be organized to answer questions and test hypotheses across and within the total sample based on: grade levels (Grades 8, 9, 10), subject areas (seven subjects), subjects within grades (eighteen subjects), and textbooks (thirty-seven prescribed texts).

Word-types are identified and features such as graphic characters and relative frequency of occurrence of individual words indicated. The repeat rate frequency of words is indicated with frequency counts and comparisons based on occurrence of word characteristics of the printed materials. Sentence length characteristics are described and compared across the grade levels and subject areas. Finally, a decision theory, or model for an elimination technique, is proposed as an aid in identifying the most significant content vocabulary in word lists derived from samples based on subject area texts.

The study is organized into nine tasks. Tasks 1 to 4 were designed to organize the input data into a total Corpus and sixty-five other corpora and develop necessary word lists and summary tables. Tasks 5 and 6 were designed to produce the descriptive and comparative statistics for the Corpus and the various corpora. Tasks 7, 8, and 9 were developed to produce analyses of selected linguistic features of the Corpus and the corpora. The nine tasks with their related questions and hypotheses follow.

Task 1. Develop a Corpus to represent natural language text based on instructional materials prescribed for use in the subject areas of British Columbia junior secondary grades.

Task 2. Organize the Corpus of materials for computer in-put and manipulation.

Task 3. Generate two volumes of the Corpus: one organized by grade levels and one organized by subject-areas, each with a descriptive index.

Task 4. Organize the samples into word lists for the Corpus, the grade corpora (3), the subject corpora (7), the subject within grade corpora (18), and the textbook corpora (37).

4.1 For each of the above, provide an alphabetical and a rank order (descending frequency) listing of word-types to give the following information.

4.11 The frequency of occurrence of each word-type.

4.12 The cumulative percentage frequency of each word-type.

4.13 The relative frequency of occurrence of each word-type per 1,000 tokens.

4.14 The descriptive statistics for the rank order lists of the Corpus and corpora including: X, FX, SUM FX, FX * X, SUM FX * X, CUM % FX * X . (A full explanation of these terms is given in Chapter III).

4.2 Construct two summary tables for each of the sixty-six word lists, indicating the word frequency figures in descending order (highest frequency first) and in ascending order (*hapax legomena* first).

Task 5. Generate comparative and statistical analyses based on the lexical characteristics of the Corpus and the corpora and data produced in Tasks 1 through 4.

5.1 What are the lexical characteristics of the Corpus; the Grade 8, 9 and 10 corpora; each of the seven subject area corpora across Grades 8, 9 and 10; each of the corpora for subjects within Grades 8, 9 and 10; and each of the thirty-seven textbook corpora in terms of: total number of graphic characters, average number of graphic characters, tokens, and discrete word-types?

5.2 What are the characteristics, in terms of repeat-rate frequency of words (Yule's K), for the Corpus and corpora defined in 5.1 above ?

Task 6. Generate comparative and statistical analyses based on sentences and sentence lengths for the Corpus, the corpora, and the data produced in Tasks 1 through 4.

6.1 What are the sentence-length characteristics of the Corpus: the Grade 8, 9 and 10 corpora: each of the seven subject area corpora across Grades 8, 9 and 10:, each of the corpora for subjects within Grades 8, 9 and 10; and each of the thirty-seven textbook corpora in terms of: average number of sentences; mean, median and modal sentence length in words; standard deviation, coefficient of variation, average number of sentences, and Pearson's skew factor for sentence lengths?

6.2 Produce a set of graphs to illustrate each of the sixty-six sentence length distributions for the Corpus and corpora defined in 6.1 above.

6.3 What are the characteristics, in terms of repeat rate frequency of sentence-lengths (Yule's K), for the Corpus and corpora defined in 6.1 above ?

Task 7. Generate comparative and statistical analyses of the distribution of the 100 most frequently occurring word-types of the Corpus across the three grade levels, the seven subject areas, and the eighteen subject areas within the three grade levels.

7.1 Test the following null hypotheses.

Hypothesis 1. There are no significant differences in the actual distribution of the 100 most frequent word-types of the Corpus when compared to the expected distribution of each word-type for:

Hypothesis 1.1 the three grade levels of the Corpus,

Hypothesis 1.2 the seven subject areas of the Corpus,

Hypothesis 1.3 the subject areas within Grade 8,

Hypothesis 1.4 the subject areas within Grade 9,

Hypothesis 1.5 the subject areas within Grade 10.

7.2 Investigate and describe the number of word-types which differ significantly in their distribution across each of the areas tested in 7.1.

Task 8. Do the sentence length distributions of the seven subject areas differ from the sentence length distribution of the Corpus ? This task involves testing the following null hypotheses.

Hypothesis 2. There are no significant differences in the actual distribution of short, average, and long sentences when compared to the expected distribution of each of the sentence lengths for:

Hypothesis 2.1 the three grade levels of the Corpus,

Hypothesis 2.2 the seven subject areas of the Corpus,

Hypothesis 2.3 the subject area corpora within Grade 8,

Hypothesis 2.4 the subject area corpora within Grade 9,

Hypothesis 2.5 the subject area corpora within Grade 10.

Task 9. Develop an "elimination technique" for selecting the most significant content words in a word list using the ranked frequency lists developed for the Corpus, the three grade level corpora, and the seven subject area corpora.

9.1 Produce a set of graphs to illustrate the word frequency by rank of the Corpus, the three grade level corpora, and the seven subject- area corpora.

9.2 What is the effect of eliminating the highest frequency words and the lowest frequency words from the total spectrum of words for each of the corpora stated in 9.1?

9.3 Can the residual of words remaining after eliminating the high and low frequency words described in 9.2 serve as a pool for selecting the most useful content words for the Corpus, the three grade level corpora, and the seven subject -area corpora through analyses based on relative frequency of occurrence and subjective criteria?

SIGNIFICANCE OF THE STUDY

The linguistic components of written discourse influence the comprehension difficulty of printed material. For example, longer sentences generally complicate the arrangement of words and make greater demands upon memory in reading than do shorter sentences. Redundancy in terms of word and sentence repetition is also considered to influence the reader in deriving information. Furthermore, comparisons of the relative frequency of occurrence of discrete word-types in different types of discourse has significance for learning and teaching, since the more a word is used, the greater the probability the reader has had an opportunity to come in contact with it.

The vocabulary loading of material, in terms of lexical and structural word-types, is another factor that influences comprehension. Even word length, measured both phonologically and graphologically in syllables or letters, can be predictive of comprehension difficulty. Increasingly, recent language research related to such factors is sharpening, both in design and quantity, and providing results with implications for teaching and further research. Now that the advent of computer technology has greatly minimized labor, such language enquiry has been stimulated by facilitating research based on large bodies of material enabling multiple comparisons across diverse sub-components.

This study will provide, first of all, a useful pool of written language samples and extensive information about the specific Corpus of materials serving as the in-put data base. In addition, programs and models will be produced which are generalizable to other idiosyncratic populations of materials and in turn can become useful tools in raising and answering further questions. The study derives its major significance from several unique features.

The study represents the first extensive analysis of the lexical characteristics of English language instructional materials prescribed for use in the secondary schools of a Canadian province. The vocabulary lists emanating from the study reflect the demands of real reading materials being used by Canadian students in the 1970's as distinct from most of the dated word lists currently in use.

The data generated by the study could have significance in developing guidelines for authors of school instructional materials and teachers using those materials. The word lists would supply writers with information they need to meet the needs and the capabilities of students, particulary those of limited reading ability. The data have value in planning instruction for both native speakers of English and students coping with English as a second language.

Data from the study would permit a number of correlational analyses to be made with existing word lists and word-graded reading tests now in extensive use throughout Canada. Researchers in related disciplines could make use of the word lists without having to rely on data from outmoded or foreign sources. The results provide a readily accessible, fundamental compilation that can be consulted when needed.

The samples obtained from the study and the related word and sentence statistics could be used to further research into the readability of secondary instructional materials and as input in the development of both standardized and informal reading tests for placement, evaluation of student progress and estimation of program effectiveness.

Improved teaching methods to facilitate instruction in reading comprehension could be a vitally important outcome of the study. Significant differences between the basic language characteristics of the subject-areas would emphasize the need to develop alternative teaching procedures and the utilization of instructional materials geared to the unique word and sentence demands of each subject-area. The results could provide data which would aid greatly in determining to what extent instructional materials reflect vocabulary that is within reach of students and in the identification of words of special importance needing special attention in teaching.

The study has potential impact beyond its specific findings, however. The model designed for the study makes extensive use of natural-language computer technology and could readily be adapted to facilitate much larger studies, or conversely the model could be used to examine very small units of material. The computer programs generated by the study could be applied in the analysis of other idiosyncratic populations of printed materials. Finally, computer technology was used extensively to produce the dissertation itself, format and print word lists, tables and graphs from raw data, and to develop, edit and produce the final printed copy. Thus the study could serve as a representative model in developing other research projects involved with the processing of natural language text.

DEFINITION OF TERMS

For the purposes of the study a number of definitions were developed.

Character
: A letter, digit, or other symbol that is used to organize, control, or represent data.

Coefficient of Variation
: A method of measuring the rate at which sentence types move away from the mean.

Computing Centre Dollar, CC$
: Used in accounting by the Computer Centre. A CC$ represents an amount of computing resources which costs the University of British Columbia $1.00 to provide.

Conversational Terminal
: A typewriter-like device which enables a user to communicate with MTS.

Corpus
: The total body of 235,107 tokens of natural language text based on the 469, five hundred word samples across thirty-seven textbooks prescribed for use in the subject areas of Grades 8, 9 and 10 in British Columbia.

Disk
: A computer storage device used in MTS for line and sequential file storage, batch queue storage, and paging.

<u>File</u>
: Used with MTS to refer to collections of related information residing on direct access devices.

<u>Magnetic Tape</u>
: A storage medium which permits the recording of data as a series of magnetized spots.

<u>MTS</u>
: The Michigan Terminal System designed to run on an IBM model computer.

<u>Pearson's Skew Factor</u>
: A method of measuring the skewness of a distribution curve.

<u>Sentence</u>
: A number of tokens, the first beginning with a capital letter and the last ending with a period, question mark, or an exclamation mark, followed by a blank space or a pair of quotation marks.

<u>Token</u>
: An individual occurrence of a word-type.

<u>Word</u>
: A continuous string of characters bounded left by a blank space and delimited by a blank space or one of the following characters '-( )";:,?@/$#+%=!_¢.

<u>Word Type</u>
: A "distinct word" representing a set of identical individual words (tokens).

<u>Yule's Characteristic K</u>
: A method of determining the rate of repetition (word-types or sentences) in a passage of print.

LIMITATIONS

There are three main limitations to the findings of this study.

1. The study is restricted to the use of thirty-seven "A" issue English language textbooks prescribed for use in Grades 8, 9, and 10 in British Columbia secondary schools during 1972-73.

Because of the size of the undertaking, not all of the content available was used. Instead, a sampling of between 30 - 40 percent of the prose selections was made.

2. No attempt is made to analyze all of the linguistic features of the material used in the study. The main focus is on the analysis of lexical characteristics and sentence forms.

3. The study is limited by the accuracy of the various computer programs which were developed specifically for the project, as well as by the accuracy of keypunching and editing procedures employed in data preparation. A Pilot Study was utilized to validate procedures and programs and minimize errors as much as possible.

OVERVIEW OF THE STUDY

The study has three major aspects: 1) the selection and treatment of the text materials used for computer input; 2) the development of computer programs needed to generate the word-lists and other related statistics; 3) the analysis and comparison of the computer generated data in relation to the questions raised and hypotheses stated.

Chapter II presents the review of literature and the conceptual framework for the study. The design and methodology of the investigation involving the nine tasks is outlined in Chapter III. Chapter IV presents the analysis of the data and the findings of the study. Finally, Chapter V presents a summary of the results, the conclusions for the investigation, and suggests a number of implications for reading instruction and future research.

CHAPTER II

REVIEW OF THE LITERATURE AND CONCEPTUAL FRAMEWORK

INTRODUCTION

This study is concerned with the development, description, comparison, and analysis of a representative sample of printed instructional material used in secondary grades. There have been few reported studies in this area and therefore it has been necessary to make use of research at other educational levels in order to construct the conceptual framework for the present investigation. Some of the studies mentioned are based on empirical research and others report the results of descriptive research.

Although the major aspect of comprehension in reading is concerned with the full relationships of phonology, syntax, and semantics, the use of printed discourse requires the reader to deal first with words. In recent years attention has been given once again to the development of word lists for the analysis of printed materials and several new word lists have been developed (Kucera-Francis, 1967; Carroll et al, 1971; Harris and Jacobson, 1972).

A recent innovation in the analysis of natural language text has been the use of computer technology. Alford (1971) emphasized that only a computer could handle the vast complexities of modern day techniques in language research. Harris and Jacobson (1972) outlined many of the advantages which a computerized system of word analysis offered, especially in comparative studies of printed materials. Information gained from research which makes use of a massive corpus of language would enable educators to select and modify instructional materials to meet the reading needs of individual students.

Research into the readability of printed materials has continued to stress the importance of word and sentence characteristics as two of the main factors in determining reading difficulty (Fry, 1968; Mclaughlan, 1969; Guthrie, 1970). Investigators using the Cloze technique maintain that having to replace randomly deleted words in passages selected as representative of print materials constitutes the best means currently available for measuring the comprehensibilities of printed prose (Bormuth, 1969; Ramanauskas, 1972). The lexical and functional aspects of the deleted words have implications related to the comprehension of print materials.

Finally an area of research in the linguistic analysis of print materials which appears to have potential concerns the development of techniques to identify the 'significant' body of words and sentences which can be used to summarize the content of a passage of material.

The purpose of this chapter is to organize a conceptual framework and make a selective review of studies that relate to these areas, including: word lists and their role in reading research; computer technology and language research; the readability of printed materials; and techniques useful in identifying the "significant" content in a body of print material.

WORD LISTS AND THEIR ROLE IN READING RESEARCH

This section deals with three major topics: the development of word lists; word lists and content materials; and word lists and readability. The first topic will include computer generated word lists to present an up-to-date outline but the use of computer technology in language research will be presented in a later section. Similarly, the role of word lists in readability research is included under the third topic, but a fuller treatment of readability will be presented in part four.

The Development of Word Lists

Extensive studies have been made of the vocabulary used in printed materials in the U.S.A since the 10,000 words listed in Thorndike's, *The Teacher's Word Book* were published in 1921. Thorndike's study, which had a great impact on educational research, made use of over four and a half million words of running prose taken from a variety of sources including children's literature, elementary school texts, commercial materials, and the Bible. Ten years later, Thorndike (1931) added another 10,000 words to the frequency lists and then collaborated with Lorge to produce a much more diverse sampling of book and magazine content (Thorndike and Lorge, 1944). The pioneer work of Thorndike and Lorge had great educational significance because there was a real need to have school instructional materials based on language which had a high functional frequency. However, the point has been made that the additional 10,000 words used in the Thorndike-Lorge list were compiled mainly from adult materials (Harris and Jacobson, 1972).

During this period a number of other researchers were constructing word lists based mainly on the language considered most common to children. Pressey (1924) compiled special vocabulary lists in fifteen school subjects in an attempt to isolate specific areas of emphasis in language usage. Gates (1926) developed a 1,500 word reading vocabulary for the primary grades by selecting from 2,500 of the highest frequency words in Thorndike's initial work, 1,000 of the most frequent words in a series of children's readers and 1,000 of the words most frequently spoken by young children. The Gates' word list had considerable influence on the vocabulary used in primary grade reading textbooks. Horn (1926) published an adult vocabulary list of 10,000 words considered to be basic for written expression and made it possible to compare this mode with reading and speaking vocabularies.

Another major development about this time was the

International Kindergarten List of 2,596 words considered most widely known by kindergarten children (West, 1928). In 1931, a list of 769 easy words which were common to both the International Kindergarten List and the first 1,000 words of the Thorndike list, was produced by Dale. The following year, the results of a complex study were presented, designed to assess which words in the English language were used most often and how other variables in the language influenced their use (Faucett and Maki, 1932). A few years later, Buckingham and Dolch (1936) developed a word list based on the word knowledge of children in Grades 2 to 6. About the same time Dolch (1936) compiled a list of 220 words by selecting 193 words which were common to the most frequent words taken from three sources: a list of 2,596 words common to preschooler's vocabularies; the Gates' Primary Word List of 1,811 words judged important in children's reading; and a list of 453 words taken from a number of primers and first grade readers.

The first major undertaking to develop a knowledge of the basic English vocabulary used in Canadian elementary schools was started in 1945 (Stothers, Jackson and Minkler, 1947). The authors pointed to the complete reliance by Canadian educators on word lists constructed in the U.S.A. They stressed that very little attention had been paid to the nature of the vocabulary in Canadian textbooks or to the role of vocabulary development as a distinctive reading skill. The method used in the study was to review a number of surveys carried out between 1921-1945 and to also assess the uncommon vocabulary of readers used in Ontario. Three word lists were then prepared: for Grades 1 and

2, Grades 3 and 4, and Grades 5 and 6 respectively. The lists were next examined by students and teachers in an attempt to check content validity. Finally a total of 5,764 words distributed across Grades 1 to 6 was presented

In 1945, a basic vocabulary for elementary school children in the U.S.A was developed by Rinsland and in 1949, the first of a series of core vocabulary lists were constructed by the Educational Developmental Laboratories (EDL). The EDL word lists were designed to facilitate the preparation of basic and supplementary reading materials and to serve as a guide for teachers and students regarding the vocabulary load of books. In addition, the basic core vocabulary was suggested for use in the development of readability levels of reading materials (Taylor, 1949).

In the initial EDL study, 150 sources were investigated. Revisions followed in 1951 and 1955, and in 1968, an additional nine basal readers were added to the survey. The primary grades' lists were based on basal readers and at the intermediate level a combination of pupils' knowledge of the word checked against the Rinsland (1945) list and the word's frequency measured against the frequency (G listing) of the Thorndike-Lorge (1944) list, determined the inclusion of a word. For Grades 7 and 8, the words from Grades 4 to 6 were rechecked against the Thorndike-Lorge (1944) and the Rinsland (1945) lists and added if their frequency warranted it. The remainder of the words were taken from the Thorndike-Lorge and the Rinsland lists. Finally the core vocabulary for Grades 9 to 13 was compiled by using the highest frequency words from the Thorndike-Lorge (1944) list as well as a number of other words from a bibliography of vocabulary improvement materials.

The Kucera-Francis (1967) analysis of American-English made use of computer techniques to compile a 1,014,232 word corpus which was unique at the time in that it was the only randomly selected sample of printed material published in the USA in the one calendar year. Fifteen genre were used in the Kucera-Francis study and 500 samples of approximately 2,000 words each were randomly selected across the genre. The study provided an invaluable data base for other researchers to use in investigating phonological and lexicographical aspects of written English language. However, the Kucera-Francis study was derived from adult materials and was not designed to provide grade level or subject area analyses of the material being treated.

Carroll, et al (1971) emphasized the need to learn more about the lexical characteristics of language in a massive study involving published materials frequently used by students in Grades 3 through 9. The American Heritage Intermediate Corpus or AHI Corpus, as the study was called, made use of computer techniques to generate frequency lists from over 5,000,000 words taken from some 1,000 different publications. The AHI Corpus was designed to provide a 'cultural frame of reference for judgment and comparison' which would serve as 'a reflection of the culture talking to its children' (Carroll et al, 1971). The word frequencies were listed by grade levels thus providing valuable information for teachers and writers of instructional materials. The authors also noted that word frequency data had been useful in helping to determine readability levels and the selection of texts for classroom instruction, the teaching of English as a second language, and the compilation of vocabulary lists. The AHI Corpus incorporated a number of statistical analyses of the Corpus by grade and subject area using the word frequency data but no attempt was made to examine sentence length characteristics of the material used in the study.

In 1972, Harris and Jacobson published a series of elementary reading vocabularies consisting of words which were widely used in elementary school textbooks during 1970. The study made use of fourteen series of elementary school textbooks for Grades 1 to 6, six basal reader series, plus two series of texts for each of the core subjects (English, Social Studies, Science, Mathematics). The lists included General Vocabulary lists containing common vocabulary found in basal readers and content textbooks, a Core List of words found in three of the six basal readers, and an Additional List made up of words which appear in four or more of the fourteen series of books used. A Core List and an Additional List were also included for each of the basal reader levels (Preprimer through Grade 6). The authors stated that their lists provided the basis for a number of comparative analyses to be made between word lists, including elements such as content, obsolescence, levels of difficulty, number and length of words, word frequency, and aspects of word construction such as singular - plural (Harris and Jacobson, 1972).

A summary of the most widely known word lists developed between 1921 - 1972 is presented in TABLE I .

TABLE I

A SUMMARY OF WORD LISTS : 1921-1972

Author	Year	Description
Thorndike	1921	The Teacher's Word Book contained 10,000 words taken from printed materials in the U.S.A.
Gates	1926	A Reading Vocabulary for the Primary Grades contained 1,500 words for Grades 1, 2, and 3.
Horn	1926	A vocabulary stated to be basic for written expression.
Thorndike	1931	Another 10,000 words added to the 1921 list.
Dale	1931	A list of 769 easy words which were common to the International Kindergarten List and the first 1,000 words of the Thorndike List.
Buckingham & Dolch	1936	A word list based on vocabularies of children in Grades 2 to 6.
Dolch	1936	A basic sight vocabulary of 220 words.
Thorndike & Lorge	1944	A much more diverse sampling of book and magazine content in the U.S.A. 30,000 words in the list.
Rinsland	1945	A Basic Vocabulary of Elementary School Children. Illustrated the frequencies of 14,571 words taken from an analysis of 200,000 written papers.
Stothers, Jackson, & Minkler	1947	The first major undertaking of produce a Canadian word-list for Grades 1-6. A total of 5,764 words used.
Taylor, Frackenpohl & White	1949 revised in 1951 & 1955	A series of core vocabularies developed by the E.D.L.

TABLE I (CONT.)

A SUMMARY OF WORD LISTS : 1921-1972

Kucera & Francis	1967	An analysis of American-English adult materials using computer techniques to generate a corpus of 1,014,232 words.
Taylor, Frackenpohl & White	1968	An additional nine basal readers were added to the 1955 revision. Lists at the primary, intermediate, and secondary levels were provided.
Carroll et al	1971	The American Heritage Word Frequency Book. A computer-generated analysis of over 5,000,000 words taken from 1,000 different publications used in Grades 3 to 9
Harris & Jacobson	1972	Basic Elementary Reading Vocabularies. A set of word lists at the elementary level develcped by computer techniques.

Word Lists and Content Materials

Studies concerned with the vocabulary content of printed materials have often followed the development of frequency word lists. Between 1925-1945 a number of researchers investigated the relationship between the vocabulary used in instructional materials in the content areas and the most common words reported in frequency word lists (Powers, 1925; Patty and Painter, 1931; Fries and Traven, 1940).

In 1952, Malsbary measured the understanding that high school students had of business and economic terms selected from a variety of newspapers, journals, and newscasts. He found that there was some relationship between student understanding and the frequency of the item. Malsbary also reported that seventy-nine of the items were understood by only 50 percent of the students.

Kyte (1953) conducted a study to determine the core vocabulary required for various instructional programs. He used the 500 most common words from each of Horn's 1926 list, the Thorndike-Lorge 1944 list and the Rinsland 1945 list. A final list of 663 words was presented.

The continued reliance of educational researchers on word lists compiled several decades earlier was reflected in a series of studies made during the early 1960's. Traxler (1963) developed two forms of a fifty-item vocabulary test for high school students and college freshmen by randomly selecting items from the 10,000th to the 20,000th word of the Thorndike-Lorge word list. Another research project compared the frequency of selected structure words found in children's and adults' written expression. The structure words were taken from Rinsland's 1945 list, Dewey's "Relative Frequency of English Speech Sounds," (1923) and from Horn's 1926 work (Card and McDavid, 1965). The following year another study analysed the frequency bias of the 122 most commonly used English words as determined in a number of lists including Dewey's and Rinsland's. The results of this study showed that structure words in English formed a typical corpus (Card and McDavid 1966).

In 1967, Jacobs compared the 1926 Buckingham-Dolch word list results with a study carried out in Oregon. He found that free-association vocabulary elicited from children in Grades 2 through 6 differed significantly from that reported in the original study. An interesting point to note is that although more children knew the same word in 1966, they also knew fewer words than their predecessors.

In 1971, Johnson made an examination of the Dolch (1936) basic sight word list and its relationship to the Kucera-Francis study. He stated that 82 of the 220 words listed by Dolch were not among the 220 most frequent words in the Kucera-Francis Corpus. Other discrepancies were reported and Johnson concluded that the original Dolch list was no longer suitable as a measuring instrument for the vocabulary content of instructional materials in the 1970's.

The need for extensive, analytical studies into the nature of instructional materials currently used in Canadian schools is an implication from the preceding discussion. Such studies would present a description of the language composition of reading materials used in Canadian education in the 1970's.

Word Lists and Readability

Word lists have been used extensively in the development of readability formulas. Lively and Pressey (1923) used the Thorndike list to give a 'weighting' to materials they had selected from elementary basal readers and college science textbooks. A number of researchers used words that were not included in the Thorndike list as a variable in their work into readability (Vogel and Washbourne, 1928; Washbourne and

Morphett, 1938; Jacobson, 1961).

Gray and Leary (1935) used the 1931 word list developed by Dale in their readability formula as did Lorge (1944) and Spache (1953). Spache later made use of the Stone (1956) revision of Dale's list in his formula. In 1948, Dale and Chall used the Dale List of 3,000 words as a variable in their readability formula. A later word list by Botel (1962) was also used in readability research.

In recent years, work into the readability of print materials has made more use of language variables other than word frequency. This aspect of readability is discussed later in the chapter.

COMPUTER TECHNOLOGY IN LANGUAGE RESEARCH

In recent years there has been a growing interest in the use of computer techniques to help compile and analyze natural language samples. The studies have concentrated in two main areas: the analysis of materials in specialized areas such as library science, information science, and foreign languages; and the analysis of educational, instructional materials.

A study which generated a computer-based general frequency word-list in German designed for use at the college level, indicated that over 30 percent of the original sample text had not been covered by previously developed word-lists (Siliakus, 1967). The author pointed out that although most of the untreated words were very low frequency items, there were also numerous high frequency proper nouns and cognates found in the portions of the text not covered. This study thus emphasized the very thorough analysis of language possible with the aid of a computer.

A later study by Johnson (1972) with Russian language materials made use of a set of frequency groups and an algorithm for the implementation of a frequency identification and marking procedure on an IBM 360 computer.

The work of Fuellhart and Weeks (1968) examined some twenty-three lexical resources in information science. This analysis, which made use of the IBM 360 Model 40 computer, was successful in quantifying the terminology and establishing the frequency of occurrence of main concepts. However, the study also showed that there was no formal structure for the discipline of information science and that the materials examined tended to reflect the opinions of the authors about the nature of the structure.

Austin (1969) conducted an investigation into the authenticity of a piece of English literature by using a computer assisted technique for stylistic discrimination. The Austin study was important in that it illustrated how frequency lists of words and other pertinent linguistic variables could be used to help determine authorship. Later research by Berkeley (1972) showed that the computer could be used to help isolate difficult terminology in a specific discipline. The computer scanned a chapter of a Navy training manual consisting of 9,800 words and classified words of two syllables or longer as either "assumed audience vocabulary" (words the audience would be expected to know), or difficult words needing further clarification. The computer scanning technique described by Berkeley made use of a previously defined lexicon and is a procedure which has important implications for future language research.

The study by Kucera and Francis (1967) represented the first attempt to computer-generate a general word-list for use in educational research through the manipulation of a massive (over 1,000,000 words) corpus. Since the Kucera-Francis work, which dealt with adult materials, researchers have been developing computer techniques to aid them in their investigation of instructional materials at various grade levels.

A study by Cronnell (1971) developed a lexicon of 9,000 words which were taken from materials used in kindergarten to Grade 3. With the use of a computer, the 9,000 words were systematically arranged both by order of word length and by the introduction of vowels in unstressed syllables. The study was designed to aid the investigation of the spelling-to-sound correspondences needed in beginning reading.

Harris and Jacobson (1972) compiled a number of elementary reading vocabulary lists taken from 127 books in twenty-eight series. This computer-assisted project produced three sets of lists which included a General Vocabulary set, a Technical Vocabulary set, and a Total Alphabetical List. The study generated approximately 17,000 word-types (after certain adjustments) from an original 80,000 unique words found in the 4,500,000 running words treated. The authors gave an excellent description of the procedures they followed throughout the investigation, including valuable information on the types of programs which were developed for use with the Burroughs B5500 computer. However, the study did not attempt to make statistical analyses of the material, but was designed to give a description of the language comprising elementary grade level textbooks in the 1970's and therefore provided an important basic reference in studies of word frequency.

A slightly different approach was presented by Durr (1973) who insisted that there was a need for an up-to-date vocabulary list at the primary level which concentrated on books which the students had selected for themselves. The author made use of eighty library books which were frequently chosen by children for recreational reading. Over 100,000 running words were then computer analyzed into word-types and a frequency list of word tokens. Durr concluded that there were several very important implications for the teaching of reading from his study including the need to introduce children to high frequency words early in their reading experience.

The first major study involving junior secondary materials was presented in the American Heritage Word Frequency Book and the American Heritage School Dictionary (Carroll et al, 1971). The authors stressed that the study, which computer processed over 5,000,000 words taken from books frequently used in Grades

3 through 9, was necessitated by both the types of material used in schools today, and by the rapid increase of new words in the English language. The Carroll study recognized the need to include materials at the junior secondary school level, but did not carry the investigation past the stage of analyzing word characteristics and dealt only with materials used in the U.S.A.

RESEARCH INTO THE READABILITY OF INSTRUCTIONAL MATERIALS

The various descriptive and statistical analyses made during this study concentrated on word and sentence characteristics as the two main language variables. The role of these variables in readability research is presented in two sections: the development of readability formulas; and work on the Cloze procedure.

The initial discussion dealing with readability formulas traced attempts by educational researchers to identify and then simplify various combinations of language variables thought to cause difficulty in reading comprehension.

The section on the development of the Cloze technique as an instrument of readability analysis concentrated on more recent attempts by researchers to gain an understanding of the syntactic and semantic relationship of language in the comprehension process.

Readability Formulas

Important language variables

Works by Chall (1958) and Klare (1963) summarized early research into readability and pointed to the need for greater expertise in the design of research studies, understanding of linguistic variables involved, and analysis of results in future research. In a study of sixty-six secondary school literature texts in use in the U.S.A, Aukerman (1965) listed five language variables which he maintained helped account for a book's readability level. Aukerman's variables were sentence length and complexity which he classed as mechanical complexity, and the incidence of verbals, word difficulty, and abstraction which he termed verbal complexity. Aukerman then constructed a table which he claimed listed the relative reading difficulty of each book based on five, 500 word samples. He also stressed that his findings were only tentative and that empirical evidence would have to wait until he had had an opportunity to engage in further research.

Klare (1966) explained that earlier work by Coleman and Bormuth had shown the value of counting letters per word as a measure of passage difficulty. Other aspects of words such as morphological complexity, the number of syllables originating in Latin, abstractness, and frequency have been investigated (Bormuth, 1967). Until recently the latter variable was not considered very important in readability analysis. However, Klare's (1968) research led him to believe that word frequency may in fact encompass most of the other variables.

Coleman (1968) outlined a number of experiments which had studied grammatical relations and readability. He stated that in most cases when a sentence is rewritten to make it more readable, it usually undergoes a grammatical transformation in some form. Coleman illustrated this by pointing out that much prose is abstract simply because the writer chose to use certain derivatives of verbs (e.g. abstract noun nominalizations from active verbs - "operation" from "operate"). Coleman concluded his article by suggesting that further research into the improved readability of instructional materials could lead to a greater awareness of the value of transformational grammar.

Rosenshine (1968) presented a description of three experiments in horizontal readability where similar passages were compared according to the cognitive similarity of their words and phrases. The findings of this study suggested several language variables which affected readability, namely indeterminate qualifiers and probability words which caused vagueness, and the omission of irrelevant sentences from the passage. Bormuth (1968) pointed out that recent research into the readability of written instructional materials had attempted to explain correlations between language and reading difficulty through a more detailed examination of the psychological processes involved.

Several recent studies into the inherent difficulties of various grammatical measures and their contribution to readability have resulted in either sentence length or word difficulty being cited as the most important variables.

MacGinitie and Tretiak (1969) used Yngve's phrase structure measurement and Allen's "sector analysis" on eighty selected passages and compared the results to the Lorge Readability formula applied to tests based on the same passages. Sentence length emerged as the variable most closely correlated with test scores.

In an experiment to investigate the learnability (new learning from a passage) as well as the readability of text books, Guthrie (1970) used eighteen linguistic variables, including sentence length, word difficulty, parts of speech, grammatical transformations, and certain other stimulus dimensions such as word familiarity. Guthrie reported that sentence length and word difficulty were the best predictors of learnability as well as readability. He supported his findings by stating that sentence length was found to correlate .842 with Cloze gain scores, while word difficulty correlated .815 with multiple choice gain scores.

Recent readability formulas

Early attempts to measure the readability levels of materials resulted in instruments which required considerable time and effort to apply. Fry (1968) developed a readability formula which used sentence length and syllables as the two language variables. Fry's formula was relatively easy to apply and correlated highly with a number of existing readability formulae.

The following year an even more simplistic and purportedly more valid readability formula entitled, "SMOG Grading," appeared. McLaughlin (1969), the creator of "SMOG," explained that after considerable research into the problem, he had concluded that polysyllabic words and sentence length were the most predictive linguistic variables to use in determining the reading difficulty of materials. McLaughlin explained his decision by pointing out that in his doctoral dissertation some three years earlier he had shown that words and sentences were respectively, the best measures of semantic and syntactic characteristics of reading difficulty. By noting that semantic and syntactic variables interact, McLaughlin claimed he was able to reduce his formula to a mere counting of the number of polysyllabic words in three sets of ten consecutive sentences, finding the square root of the number obtained, and adding three. He then gave a detailed account of the validity of his instrument and emphasized that it gave a measure of complete understanding of the material in contrast to other formulas which stated a 'general understanding' only. For this reason, McLaughlin concluded, the "SMOG Grading" would usually be several grades higher than other readability formulas in common use.

New trends in research

In a well designed series of studies, Bormuth (1969) illustrated just how far research into readability was from achieving its objective and stated, "It had been anticipated that these analyses would simplify the problem of constructing the theory of the comprehension of language. The failure to realize this expectation was spectacular". A main objective of Bormuth's studies was to isolate linguistic features of printed passages and determine which features stood in causal relation to comprehension difficulties. The materials consisted of 330 passages, drawn from ten subject areas, representative of Grades 1 through 12. The 169 linguistic variables included eight vocabulary variables (factors such as letters per syllable, syllables per word, frequency of lexical and structural words and the like); fifty variables based on syntactic structures (including factors which might indicate how the types or numbers of structures a sentence contains might influence comprehension difficulty); thirty-eight syntactic complexity variables (including structural density, transformational complexity, structural complexity, Yngve depth and syntactic length); sixty-two parts of speech variables, and eleven anaphora variables (including frequency of anaphoric structures, density of of anaphora, and the time interval between occurrence of an anaphora and its antecedent).

A total of 94 of the 169 variables correlated significantly with measures of passage difficulty including 8 out of 8 vocabulary variables, 19 out of 50 syntactic structure variables, 34 out of 38 syntactic complexity variables, 25 out of 62 parts of speech variables, and 8 out of 11 anaphoric variables. It is interesting to note that all 8 vocabulary variables and 11 out of 12 of the syntactic length variables, included in syntactic complexity, correlated significantly with passage difficulty. The high number of significantly correlated variables suggested that an overwhelming number of answers might presently be given to the question, "What accounts for comprehension difficulty of printed materials?" In addition, variables not specified in the total examined in the study could relate significantly to comprehension difficulty. Some estimates place the total number of possible variables at well over 200. The 94 significantly related variables were also factor analyzed. Bormuth stated, "To summarize the results from factor analysis, then, a simple structure does not seem to underly the variables correlating with passage difficulty".

In order to facilitate valid research into the readability of printed materials, Bormuth advocated the use of very large samples of words to enable rarely occurring linguistic variables to be adequately examined. Results obtained from such studies would offer valuable guidance to educators concerned with the construction of instructional materials suitable for students at varying levels of reading ability. The use of computerized technology offers exciting possibilities in this regard in the near future.

The Cloze Procedure

The deletion of words from a passage of print materials at regular intervals ensures that both lexical and structural words will be omitted. When Cloze tests have been constructed and administered correctly, the results are said to measure the facility a student has in understanding the syntactic and semantic interrelationships of the material being read.

A description of Cloze

The Cloze technique, which was used by Ebbinghaus as early as 1897, was first developed as an instrument for measuring reading comprehension by Taylor (1953). Basically the Cloze readability procedure involved five steps:

1. The selection of passages from the material to be evaluated,

2. The deletion of every "nth" word (usually every fifth word) and the insertion of underlined blanks of a standard length,

3. The administering of the mutilated text to students who had not read the original work,

4. The instruction to students to write in the blank spaces the words they thought had been deleted,

5. The marking of correct responses when identical items have been inserted (Bormuth, 1968).

Since the work of Taylor, there have been numerous studies into the application of Cloze as a means of measuring (a) comprehension, (b) readability, and (c) language variables. A survey of some of the studies pertaining to the latter category as it relates to the secondary school level will be presented in this section of the chapter. Much more comprehensive treatments of the Cloze procedure have been organized by Rankin (1965), Weintraub (1968), Potter (1968), Bickley, Ellington and Bickley (1970), Jongsma (1971), Bormuth (1972), and Bailey (1973).

Important linguistic variables

Louthan (1965) noted that when specific words were deliberately deleted, increased emphasis was placed on the meaning of the remaining words in context. The Cloze technique, therefore, made the reader use all of the lexical and grammatical clues inherent in the language structure. In a series of experiments, Loutham deleted a variety of linguistic variables including parts of speech and function words, and tested to see if the experimental subjects improved in reading comprehension compared to a control group who read material which had not been mutilated. The experimental group which showed the most significant gain in comprehension was the one which was reading material with certain function words (a, the, that, whose, what, his) deleted.

Another researcher attempting to discover more exact predictors of comprehension difficulties by using the Cloze technique, stated that important information could be learned about the difficulties of words, independent clauses, and sentences (Bormuth, 1966).

The vital importance of content words in language was illustrated by a study conducted by Weaver and Bickley (1967). It was found that originators of written material could recall about 85 percent of both structural and lexical deletions two days after their writing, whereas students who had read the materials only, could recall structural words as well as the producers, but could not recall lexical words. The importance of accurate information pertaining to specialized core vocabularies for each of the content areas was an obvious implication to be drawn.

Bickley, et al (1970) pointed out that research in Cloze had been conducted into the effect of sentence length on the comprehension of the reader and that short sentences were found to be more readable than long sentences.

Cloze and readability

A number of studies designed to explore the suitability of instructional materials using the Cloze technique were reported in 1968. Weintraub (1968) reviewed several studies which showed that the Cloze technique had high reliability and validity in measuring readability. It was further suggested that Cloze could offer valuable insights into aspects of the reading process. Bormuth (1968) investigated the relationship between the readability level and the amount of information gained by the reader. He stated that scores on Cloze tests did not depend entirely on the reader's prior knowledge of the material. This would suggest that the role of certain function words in the language structure was of vital importance. Bormuth's other contributions to research in the Cloze technique have aided the work into readability tremendously. His early work concentrated on the need to develop the Cloze procedure into an effective instrument to use in studies of readability. By this means, Bormuth planned to identify 'the linguistic features that serve as stimuli for the various comprehension processes' and then to move 'towards efforts to operationalize those processes in a manner that is suitable for instruction.' Thus the application of the Cloze procedure would enable a greater understanding to be gained of the causal relationship between specific linguistic variables and levels of comprehension among secondary school students.

An excellent summary of experiments using the Cloze technique to determine readability levels of materials for children and adults was presented by Potter (1968). In addition to his discussion on the technical aspects of the studies, Potter mentioned that the separate scoring of function and content words may provide valuable information for specialized purposes. Geyer (1968) tested the use of Cloze as a predictor of a student's ability to comprehend social studies content and also to determine if materials rewritten at an easier readability level would result in improved comprehension. The results of the latter aspect of the study showed that comprehension may not be significantly improved by reducing vocabulary difficulty and sentence complexity. Hater (1969) and later Kulm (1971) measured the readability of mathematical English. Kulm reported that there were at least ten language variables that had a significant effect on the readability of the material. Kulm maintained that existing readability formulas that rely on word difficulty and sentence length tc measure readability are not appropriate to use with mathematical English. The work of Houska (1971) showed that the Cloze procedure was a viable instrument to determine the readability level of instruction materials in Industrial Education at the secondary school level. An interesting approach was offered by Ramanauskas (1972) who conducted an experiment using two examples of material with identical syntactic and semantic components but with some of the sentences rearranged in the second sample. Ramanauskas argued that the readability of the second sample, as measured by readability formulas, was unchanged. That is, there were exactly the same number of sentences, words, syllables, etc. as before. It was only by using the Cloze technique that a valid measure of readability could be obtained.

DETERMINING SIGNIFICANT CONTENT MATERIAL

Words are generally considered to belong to one of two classifications: structure or function words, and lexical or referential words (Betts , 1965; Dauzat, 1968). The former words act as clues to grammatical structure (e.g. a, an, at, by, what, very) whereas the latter type have lexical meanings which are readily distinguishable from structural meanings. Fries (1952) identified some 154 structure words and categorized them into fifteen groups including auxiliary verbs, conjunctions, prepositions, relative pronouns, and determiners. However, Fries did not claim that his list was exhaustive and other writers have defined considerably more words as structure words (Lefevre, 1964; Goodman et al, 1966).

The role of function words in providing structural information was illustrated by Young (1973) who quoted a portion of Lewis Carrol's poem "Jabberwocky":

> 'Twas brillig, and the slithy toves
> Did gyre and gimble in the wabe;
> All mimsy were the borogoves,
> And the mome raths outgrabe...'

Young pointed out that the underscored structure words helped generate the ideas which were inherent in the nonsense sentences comprising the poem.

Rogers (1965) stated that although structure words were relatively few in number they were extremely important in the language because of their dense distribution. Fries (1952) found that structure words accounted for over 30 percent of a sample of 1,000 words while Kucera and Francis (1967) estimated that just under 50 percent of their megaword corpus consisted of structure words. The Kucera-Francis study showed that the frequency of structure words differed greatly across the various genre samples. The vast majority of the 100 most frequently occurring words in each of the fifteen genre examined were function words. However, the rank order of the structure words (except for "the" which was first in all cases) was noticeably affected by the type of genre in which it occurred; "if" was the second most frequent word in ten of the genre, while "and" was second in rank in five genre.

An area of research which is pertinent to this study is the automatic creation of a literature abstract derived from an analysis of words in a literary passage. Luhn (1958) outlined the methodology of "auto-abstracting" which involved determining a word-list compiled in descending order of frequency to give a "significance" factor for words, and an analysis of the relative position of the words in each sentence to determine the significance of sentences. A combination of these two measurements was then used to give a "significance" factor for sentences. The "auto-abstract" was finally compiled from the highest ranking or most significant sentences. Luhn defined the most "significant" words as being neither in the region of highest frequency (these words constituted the 'noise' in the system), nor in the area of low frequency where their rarity would negate their relevance to the subject matter. The "significant" section of words in the material would therefore occur somewhere between the two extreme points in the distribution. Luhn hypothesized that it would then be possible to determine the degree of discrimination or "resolving power" of the words making up this middle section of the distribution. The "significance" factor for sentences was arrived at by identifying the proximity of "significant" words to one another. Sentences which had the greatest number of frequently occurring different words in close proximity to each other were ranked higher and were classed as more "significant" to the subject. These sentences were then selected on the basis of their rank to form the "auto-abstract" of the excerpt or article. An obvious drawback to the system described by Luhn was the absence of intellectual decisions made by specialists in various disciplines in making a final selection of "significant" content.

A similar technique for automatically analyzing printed materials was suggested by Maron (1961) who was concerned with the automatic indexing of documents according to their subject content. Maron's thesis stated that reasonably valid predictions of the subject matter of documents could be made on the basis of statistics involving word frequency, word order, location, etc. The main difficulty concerned the selection of clue words which would be neither too rare to be valid predictors, nor belong to the 'logical' class of structure words which did not supply referential meaning for the material. Maron decided that the high frequency structure words which accounted for over 40 percent of the total occurrences in his study should be excluded because of their lack of information about the subject matter. Similarly, the high frequency lexical words were next excluded because of their lack of specificity for the subject. Next to be rejected were the words which occurred only once or twice in the corpus. The resulting 1,000 words were then listed and analyzed to determine which of these words were valid predictors of the subject content.

Although the present study did not attempt to make an intensive analysis of the data along the lines suggested by Luhn (1958) or Maron (1961), Task 9 in CHAPTER I suggests a possible strategy for further research into the analysis of print materials and the selection of significant content vocabulary. The various grade level and subject area corpora generated by this study also offer a readily available sample of materials for the purpose of developing techniques for the selection of significant content vocabulary in print sources.

SUMMARY

Numerous studies have been made into the vocabulary used in printed English language materials since the 1920's. Most of the studies have originated in the U.S.A and have concentrated on word-lists designed for use in primary and elementary schools. The development of frequency counts of words occurring in written discourse has aided other researchers in their examination of both linguistic and psychological aspects of the language. Vocabulary lists provide one of the most important factors in readability work and much of this research relies on the availability of word lists. Sources of this nature at the secondary school level have been lacking in the past.

The latest trends have seen the use of digital computers which have allowed researchers to deal with much greater and more diverse amounts of printed materials based on careful procedures of random sampling. The need remains for similar, well-designed studies to be made into instructional materials used in Canadian schools at the secondary level.

Few word lists have been developed based on representative sampling from secondary subject area materials which allow for analysis across grades and by subject area. Word lists traditionally provide data in frequency of occurrence of word-types but do not indicate repeat rate frequency or averages or variability for samples organized by grades or subject areas. In addition, only a small number of studies have reported sentence length characteristics, indicating averages and variability, or repeat rate frequency for sentence length types for samples organized by grades and subject areas. Although subjective analyses have been reported, few studies based on data from carefully selected print sources have been announced which empirically validate subjective opinions with respect to word and sentence characteristics of samples of natural language text from subject areas.

Early attempts to construct readability measures resulted in formulas that required lengthy and fairly complex computations. Later research into the most significant language variables involved in readability, isolated word difficulty (often measured as word length and word frequency) and sentence length as two important variables. Recent research into readability has emphasized the need to develop much larger samples of instructional materials and more usable word lists on which to base investigations. Also the need to look much more closely at linguistic variables which appear to have a causal relationship to comprehension difficulty is of prime importance. Again, data bases consisting of carefully selected representative samples from natural language text are necessary in furthering this type of research. Because of the numbers of samples involved and the complexity of the linguistic variables which need to be examined, data bases should be organized for computer input and processing.

The use of the Cloze technique to measure readability has gained considerable attention since the late 1950's. This procedure involves many aspects of language including lexical and structural words, grammar, and connotative features of language. Many researchers feel that in the future the Cloze technique will be able to contribute a great deal to an understanding of the syntactic and semantic functions of the basic language variables in instructional materials at the secondary level. Effective Cloze analyses are also facilitated by the availability of well organized data bases that have known word and sentence characteristics.

The development of methodology to identify significant content material in a printed passage was recommended by several researchers. The techniques have potential in the examination of word lists derived from samples of print materials from subject areas. Many studies generate word lists but little attention is given to the provision of adequate techniques for the identification of the significant content in such lists.

In summary, the focus of this study is on the identification and analysis of the lexical characteristics of a sample of print materials prescribed for use in junior secondary subject areas. The conceptual base, design and methodclogy for the study emanate from the review and analysis of selected, related literature in the four areas previously discussed. A well defined, representative, adequately stratified body of print material consisting of 500 word samples, forms the basis for the development of word lists and comparative and statistical analyses. The samples are organized to represent characteristics of the prescribed print materials across the total junior secondary curriculum, by the three grades, by the seven subject areas across grades, by the eighteen subjects within grades, and by the thirty-seven textbooks. The word and sentence data are analyzed in terms of the relative frequency of occurrence of various word-types and sentence lengths, and the empirical tests made to illustrate the pattern in frequency of occurrence of the most common words and a series of representative sentence lengths. A technique is proposed which serves as a model for the identification of the most significant content in word lists derived from subject area materials. Finally, computer technology is used throughout in the development, organization, comparison and analysis of the data base and in the production of the final printed copy of the dissertation itself.

CHAPTER III

THE RESEARCH DESIGN

This chapter describes the research design and methodology for the study. The study was concerned with 'present-oriented' research and a descriptive, survey approach was used to describe a specific set of phenomena in and of themselves utilizing unobtrusive measures derived from samples of natural language text. The information provides the answers to the research questions and hypotheses posed. The research method was developed to make an accurate assessment of the incidence, distribution, and relationships of the phenomena under investigation.

The research design was organized to generate the samples of natural language text, produce the Corpus of materials, develop the various word lists and generate the data necessary to accomplish the nine major tasks, answer the questions raised, and test the specific hypotheses of the study as outlined in Chapter I. A Pilot Study was first conducted to generate needed computer programs, test procedures and sharpen the methodology for the study. Following a description of the Pilot Study, the design and methodology for each of the nine major tasks are presented.

THE PILOT STUDY

Before commencing with the study it was necessary to make a trial run with a small sample of instructional materials. This procedure was utilized to determine:

(a) the time needed for keypunching a set amount of running prose (10,000 words) so that an estimate could be made of the eventual size of the data base to be used in the study,

(b) the incidence of errors in keypunching to determine whether it was necessary to have the work verified by machines,

(c) the efficiency of existing programs and the need for additional programs necessary to organize word lists, make statistical analyses, etc,

(d) the size of the samples taken from each text needed to give a valid representation of the content material,

(e) the reliability of using a random, stratified sampling technique within a textbook,

(f) the use of delimiters to determine words and sentences in the content material, and

(g) the amount of data that could be feasibly analyzed within the time and resources available.

Twenty-one samples of approximately 500 words were taken from the prescribed "B" issue textbook for Agriculture, Farmer's Shop Book (See Table II). This particular text was chosen for the Pilot Study because in the judgment of the researcher the material contained a good selection of both verbal and symbolic language likely to be found in the other content areas.

TABLE II

THE TWENTY-ONE, 500 WORD SAMPLES USED IN THE PILOT STUDY

Sample	Pages	Sample	Pages
01	14-17	12	231-233
02	34-36	13	261
03	62-64	14	274
04	78-81	15	313-316
05	109-111	16	328-329
06	129	17	353-355
07	146-147	18	375
08	161-164	19	390-391
09	189-190	20	416-417
10	216-220	21	433-437
11	223-275		

The rate of error was found to be less than one word per 500 words keypunched which suggested that machine checking was not warranted. The rate of keypunching was estimated at approximately 1,000 words per hour under ideal conditions.

As a result of the Pilot Study, the following decisions were made:

(a) A Corpus of approximately 235,000 words of running prose taken from 469 samples of 500 words each was feasible for the study.

(b) The Command Operand ")P", which was interspersed throughout the text input to signal a new paragraph, was deleted from the final frequency count of words because of its high rate of occurrence.

(c) An additional nine delimiters of a word were included to bring the total to twenty-one. These consisted of:

'-( );:,¬?@/$#%+=!_¢.

(d) The dictionary size established to deal with word-types was set at 20,000.

(e) A "Repeat Rate Frequency" table designed to illustrate the incidence of similarly occurring frequencies for both word-types and sentence lengths was included for each frequency word list and for the sentence analysis.

(f) The chi-square and Yule's Characteristic "K" statistics were tested and included in the study for both word frequency and sentence length analyses.

(g) A number of additional programs were developed to enable data to be generated in the form desired.

(h) The graphs depicting word frequency and frequency of sentence length were plotted by computer programs.

TASK 1. SELECTION OF MATERIALS

The sampling procedures were developed to provide representative lexical data for every prescribed text with sufficient quantities of natural language prose. The selection procedure consisted of two phases: an initial subjective decision to determine the number of texts and samples to be used, followed by a stratified, random sampling procedure to determine the number of samples to be selected within each text.

Works of verse or drama were not included on the grounds that they seemed to involve special linguistic problems and did not constitute the usual syntax associated with normal prose. Passages containing special coding techniques such as shorthand

and mathematics were excluded for the same reasons.

Sampling Procedures

(a) Textbooks And Samples Included. Thirty-seven "A" issue textbooks containing samples of English language prose of 500 words or over were included in the study. The total number of textbooks and samples for each content area is presented in TABLE III. Information pertaining to titles, authors and publishers of the books is listed in APPENDIX A.

TABLE III

NUMBER OF TEXTS AND SAMPLES FOR EACH GRADE LEVEL AND SUBJECT AREA

SUBJECT	GRADE 8		GRADE 9		GRADE 10		SUBJECT TOT	
	Text	Sample	Text	Sample	Text	Sample	Text	Sample
Commerce	X	X	2	25	2	16	4	41
English	2	17	4	47	2	16	8	80
Home Economics	1	22	5	76	X	X	6	98
Ind.Education	1	9	3	54	X	X	4	63
Mathematics	1	14	1	7	1	14	3	35
Science	2	20	2	24	2	31	6	75
Social Studies	2	22	2	13	2	42	6	77
Grade Totals	9	104	19	246	9	119	37	469

The thirty-seven "A" issue textbooks were selected for use because every student in each class or course of study receives a copy of the text. Other texts include those provided in sets to be shared by the students, ("B" issue); prescribed for teacher use only, ("C" issue); or allotted for special purposes, ("D" and "E" issue) and are described in the booklet, Prescribed Textbooks, 1972-73, Grades I-XII, published by the Department of Education, Province of British Columbia.

Eleven "A" issue textbooks were not included in the study. The grade level and subject area of the omitted textbooks (with the number of texts in brackets) were as follows: Grade 10 Commerce (2), Grade 8 English (1), Grade 9 English (1), Grade 10 English (2), Grade 9 Mathematics (1), Grade 10 Mathematics (1), Grade 8 Social Studies (1), Grade 9 Social Studies (1), Grade 10 Social Studies (1). The reasons for excluding the textbooks are as follows: the Commerce textbooks contained shorthand exercises; the English textbooks consisted of poetry and blank verse; the Mathematics textbooks contained mainly algebraic and geometric problems; and the Social Studies textbooks were atlases.

The textbooks used in Grade 10 Home Economics and Industrial Education were the same as those prescribed for use in Grade 9 and were included only once because the repetition of identical material would have distorted the results obtained. A Grade 9 English textbook used in Grade 10 English was not included in the latter total for the same reason.

(b) Sampling. A total of 469 samples, each of approximately 500 words, were selected from the thirty-seven "A" textbooks (see APPENDIX A). Samples of 500 words were used because research evidence suggested that there was both greater diversity of word-types from samples of this size than larger samples, and sufficient flexibility in the representation of content materials (Carroll et al, 1971).

The samples consisted of English language running prose and were randomly selected from every twenty pages throughout each text using a table of random numbers. Each sample began with the first complete sentence on the page selected and continued for approximately 500 words. Everything other than running prose was omitted, including titles, running heads, footnotes, tables, and picture captions. Two lists of the sample sizes, one in alphabetical order and one in ascending rank order, are presented in APPENDIX B. These procedures produced a total random sample of approximately 40 percent of the "A" issue instructional materials prescribed for use in the seven subject areas of Grades 8, 9, and 10 in British Columbia junior secondary schools.

TASK 2. INPUT PROCESSING, KEY PUNCHING AND EDITING

Once the sampling procedures were established, the selections were keypunched onto computer cards using the UBC FORMAT (FMT) program. FMT is a program which enables the rapid printing of materials in upper and lower case and with special characters directly on the system printer.

Input to the program was in free-form lines. The material was formatted and controlled according to <u>control</u> cards and <u>command</u> words interspersed throughout the input. The basic command operands and special operand values of the FMT program were used to organize the format of the document and allow for most instances of special arrangment (indenting, centering, underlining, etc) usually required in a formal paper. In addition, the symbol ¬ was placed after a period that did not signify the end of a sentence (e.g. Dr.¬).

Each sample was given a code number consisting of the grade level (designated by 1, 2, or 3 for Grades 8, 9, and 10); a letter signifying the subject area, Commerce(B), English(C), Home Economics(D), Industrial Education(E), Mathematics(F), Science(G), Social Studies(H); a two digit number for the order of the text; the letter "C" to represent the Corpus; and another two-digit number to distinguish the sequence of samples in each text. Thus 2B 01 C 01 designated the first sample in the first textbook listed for Grade 9 Commerce and 3H 03 C 07 represented the seventh sample in the third textbook listed for Grade 10 Social Studies.

The information on the cards was then transferred in 80 character "card-image" form to a magnetic tape and stored permanently in the computer library. The equipment used was an IBM /370 Model 168 computer, with 2 megabytes of storage, and five 9-channel 1600 bpi IBM 2401 tape drives. The computer has 14 ITEL 7330 disk units and four 1100-line-per-minute printers, plus a number of card readers and card punches.

A more detailed explanation of the 209 computer files and programs developed for processing the input samples and conducting the analyses is provided in APPENDIX C. These programs are available at the Computing Centre, University of British Columbia.

Text Corrections. Three methods of text correction were used to ensure accuracy.

1. A preliminary stage of proof-reading took place when the content of each sample had been keypunched onto IBM cards. The cards were scanned by the writer and checked against the original text. The cards were then printed as FMT output and the print-out was again scanned for obvious errors. Verification by machine means was not used because of the small incidence of errors noted in the Pilot Study and also because of the high cost of this method.

2. The second stage of editing made use of the Conversational Terminal which is an IBM 3270 Display Station consisting of a cathode-ray-tube screen (CRT) and a typewriter-like keyboard. The original input data were displayed on the CRT and scanned again for errors. Corrections were made and a revised print-out was obtained for examination. The use of the Conversational Terminal facilitated very fast proof-reading and correction of the material being processed.

3. The final stage of proofing was possible after the Corpus vocabulary had been arranged in descending order of frequency of word-types. This method of editing was by far the most efficient for it merely entailed checking the *hapax legomena* (words that occurred once) and words that had occurred twice to quickly identify obvious errors. The chance of a word

being incorrectly keypunched more than twice in different parts of the corpus was considered unlikely and a quick check confirmed this belief.

TASK 3. PRODUCTION OF THE CORPUS

Task 3 involved the use of existing programs and the development of new programs to generate two copies of the Corpus: one organized by subjects within each of the three grade levels, and the other organized by subjects across the Corpus. An index was also developed for each Corpus which listed the full description of the textbooks and the samples used in the study. The two corpora have been produced as separate volumes and are identified as CG (Corpus by Grades) and CS (Corpus by Subjects). The MTS FORMAT computer program was used tc produce the print-out of these corpora. The production of the two copies of the Corpus is illustrated in FIGURE 1. A detailed description of the computer programs and procedures followed to produce the two copies of the Corpus is presented in the document, Programmer's Guide to the Edwards' Corpus, by Allan Miller (1974).

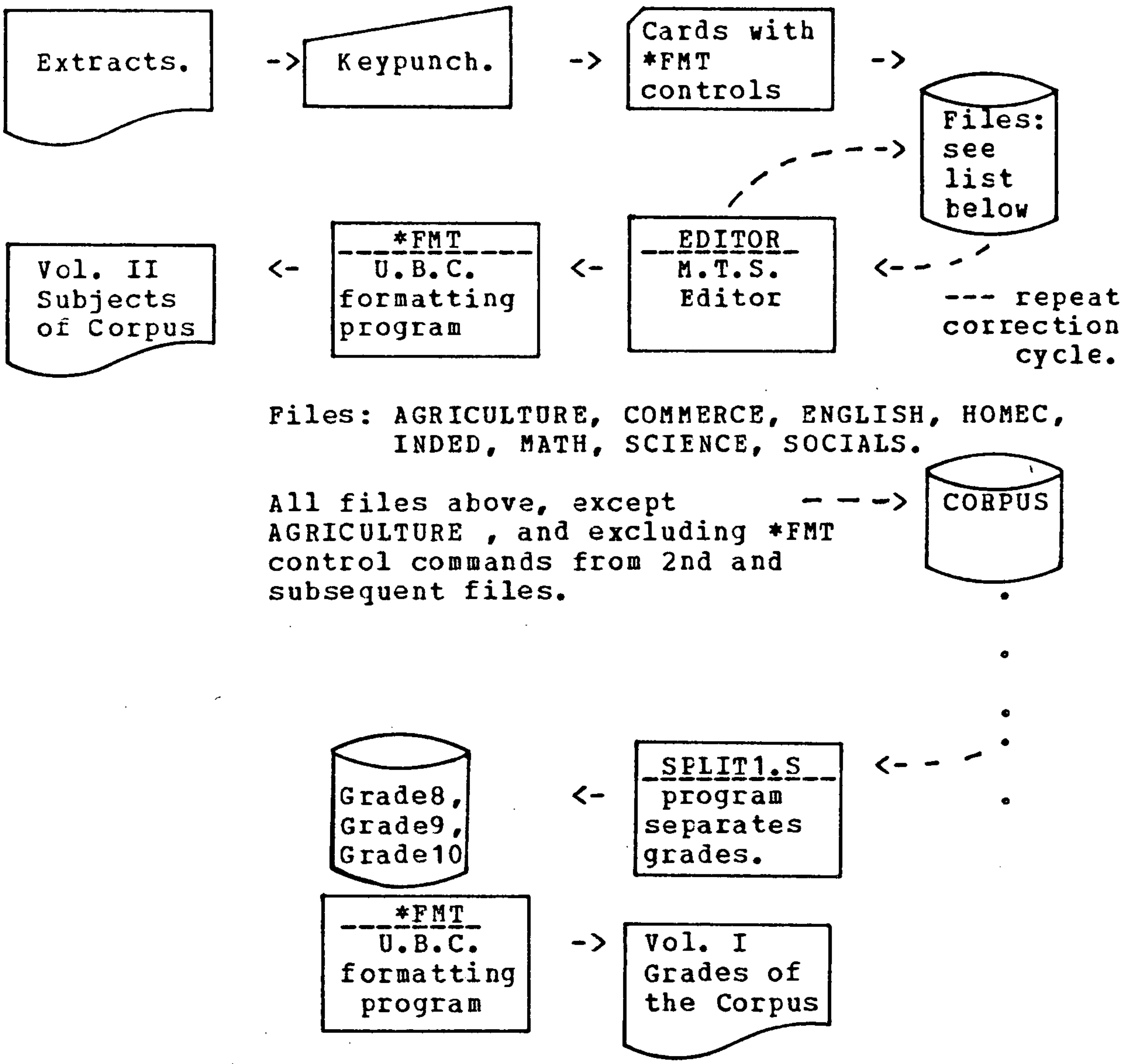


FIGURE 1

PRODUCTION OF VOLUMES C.G. AND C.S. OF THE CORPUS

TASK 4. PRODUCTION OF WORD LISTS

This task involved the generation of word lists based on various combinations of samples into distinctive corpora. Two sub-tasks were involved.

<u>Task 4.1</u>

Existing computer programs were utilized and new programs developed where needed to generate two lists based on each of the following sixty-six corpora; a) the Corpus; b) three grade corpora; c) seven subject corpora; d) eighteen subject within grade corpora; e) thirty-seven textbook corpora.

The first of the two lists is an alphabetical arrangement of word-types (See APPENDIX D). The list consists of three columns and a number placed at the top of the first column provides a running total of the word-types to that point. The first column (FREQ) indicates the relative frequency per 1000 tokens for each word-type; the second column (COUNT) states the frequency of occurrence; and the third column (WORD) lists the word-type.

The second list presents the rank-order of each word-type (See APPENDIX E). The rank list also consists of three columns similar to the alphabetical list except that the first column (FREQ) indicates the cumulative percent of the total corpus accounted for by each word-type.

<u>Task 4.2</u>

Two additional tables were included for the rank-order list which summarized the rank in (a) descending order (i.e. the word of highest frequency first and the <u>hapax legomena</u> last); and (b) ascending order (i.e. the <u>hapax legomena</u> first and the highest frequency word last). (See APPENDIX F).

The organization of the tables and the column headings are identical, except that the table which gives the descending order has an extra column (RANK) which provides the rank number of each word-type. This arrangement makes it possible to quickly locate the rank of any word in the Corpus or the various corpora by matching the frequency of a word in either the alphabetical list or the rank-order list under the column COUNT, with the same frequency in column X in the descending order table. In cases where the frequency of word-types is the same, the rank range of these words is supplied.

The column headings in the descending and ascending order tables provide the following information.

Column X

The frequency of occurrence of tokens.

Column FX

The number of word-types of the frequency X.

Column SUM FX

The sum of word-types counting from the top of the table.

Column CUM% FX

The sum of word-types as a cumulative percentage of the total number of word-types.

Column FX * X

The number of tokens accounted for by each of the word-types.

Column SUM FX * X

The number of tokens due to the cumulative total of word-types.

Column CUM% FX * X

The previous column as a percentage of the total number of tokens.

The descending and ascending order tables facilitate the rapid analysis of the information contained in the various word lists. For example, the descending order table shows that the first 100 most frequent word-types in the Corpus account for a mere 0.610 percent of the total number of word-types. However, the same 100 words-types account for 48.973 percent of the total number of tokens in the Corpus. On the other hand, the ascending order table shows that words occurring ten times or less account for 84.505 percent of the word-types but only 14.705 percent of the total number of tokens.

The word lists and accompanying tables described in Task 4 have been organized into the following five volumes:

1) Corpus, designated as C.V. (Corpus Vocabulary);

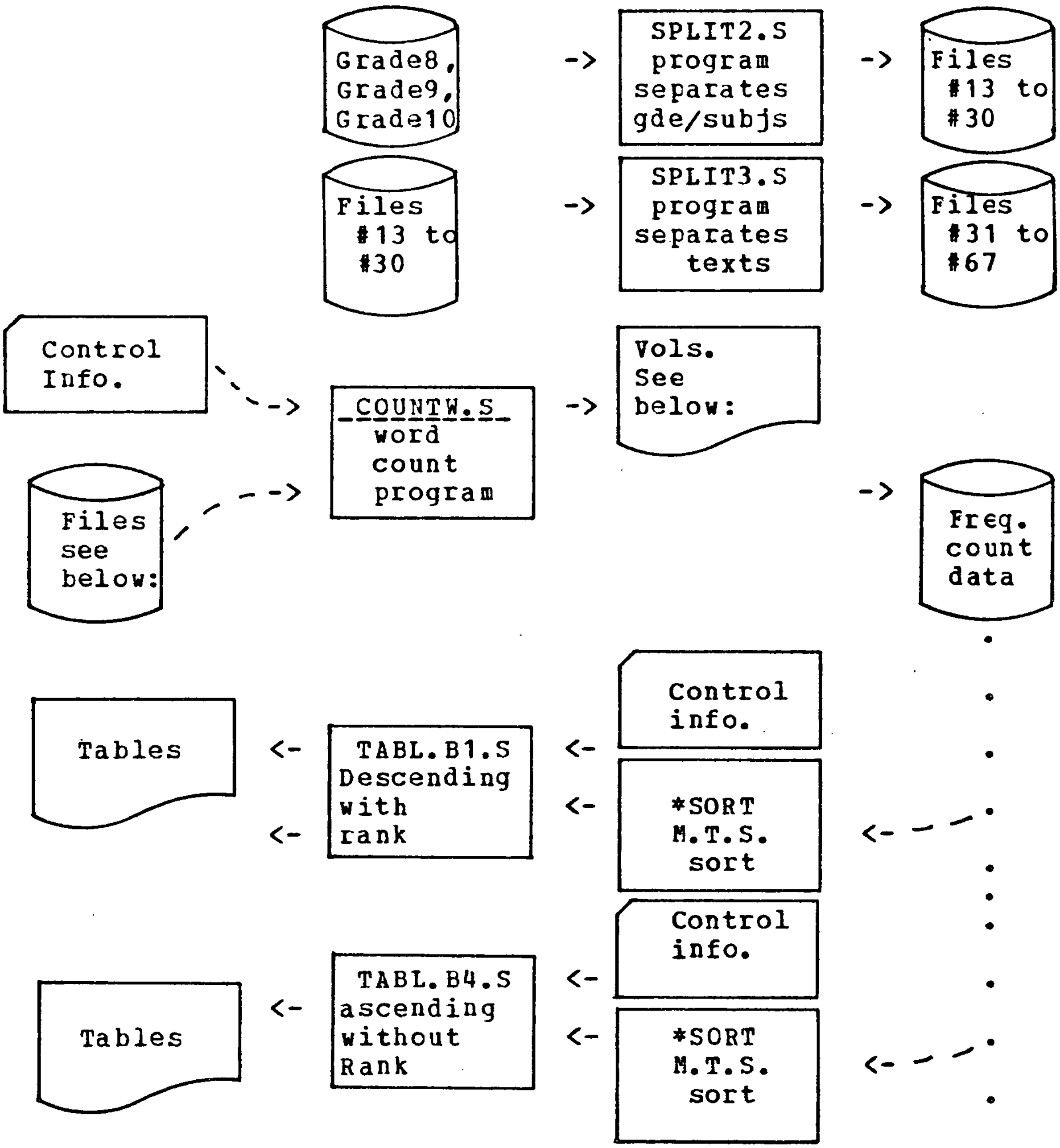
2) Grades, designated as G.V. (Grade Vocabulary);

3) Subjects, designated as S.V. (Subject Vocabulary);

4) Subjects within Grades, designated as S.G.V. (Subjects by Grade Vocabulary); and

5) Textbooks, designated as T.V. (Textbook Vocabulary).

The production of the volumes discussed in Task 4 was accomplished by making use of a number of computer programs as illustrated in FIGURE 2 .



Files: all 67 raw data files listed
in APPENDIX C with the exception
of AGRICULTURE (the Pilot Study).

Volumes: C.V., G.V., S.V.,
S.G.V., and T.V.

FIGURE 2

PRODUCTION OF WORD LISTS: VOLUMES C.V., G.V., S.V., S.G.V., AND T.V.

A complete description of the word lists and the computer programs used is available on Tape #RE0616 at the University of British Columbia Computing Centre.

TASK 5. DESCRIPTION AND ANALYSIS OF LEXICAL CHARACTERISTICS

For Task 5 existing computer programs were utilized and new programs developed where necessary to generate a number of comparative and statistical analyses for the Corpus, the three grade level corpora, the seven subject-area corpora, the eighteen subject within grade corpora, and the thirty-seven textbook corpora.

Task 5.1 was designed to determine the number of word-types, tokens, characters, and average number of characters per token for each of the following: a) the Corpus; b) three grade corpora; c) seven subject corpora; d) eighteen subject within grade corpora; and e) thirty-seven textbook corpora. Comparative summary tables were developed for this data and are presented in Chapter IV.

Task 5.2 was designed to determine the repeat-rate frequency for word-types for each of the following: a) the Corpus; b) three grade corpora; c) seven subject corpora; d) eighteen subject within grade corpora; and e) thirty-seven textbook corpora.

The repeat-rate frequency tables of word-types for each of the sixty-six corpora are included in the five volumes C.V., G.V., S.V., S.G.V., and T.V. (See Task 4.2). The first column (REPETITIONS) of each table gives the frequency of the word-type and the second column (RATE) indicates the number of word-types that have this frequency. The tables thus combine like frequencies of word-types and present different information than the basic tables of word frequencies discussed earlier.

Task 5.3 makes use of Yule's characteristic K which is a statistical parameter of a frequency distribution based on the Poisson probability law. The assumptions underlying the use of the K characteristic have been stated theoretically (Yule, 1944) and tested empirically (Kucera and Francis, 1967). In brief, the K factor is said to be independent of sample size when the samples have been collected from a large body of materials.

Formula for K:

$$K = 10{,}000 \quad \frac{S1 - S2}{S1^2}$$

where $S1 = \sum_x fx\ X$ is the first moment of the distribution about zero as origin, $S2 = \sum_x fx\ X^2$ is the second moment, and fx is the number of words occurring X times. The quantity 10,000 is introduced to avoid dealing with small decimals.

Yule's characteristic K was used to provide an indication of the concentration of vocabulary in the samples from a particular area. A large K value implies a greater use of commonly occurring vocabulary or words of high frequency of occurrence. A low K value implies that the material contains a greater proportion of rare words or words of low frequency. Summary tables of Yule's K for each of the sixty-six corpora are presented in Chapter IV.

TASK 6. DESCRIPTION AND ANALYSIS OF SENTENCE CHARACTERISTICS

A number of existing computer programs were used and others modified for use in the development of Task 6. Comparative and statistical analyses were generated, based on the sentence and sentence length characteristics of the sixty-six corpora of the study. Four major sub-tasks were involved.

Task 6.1 was designed to provide sentence length characteristics including mean sentence length, standard deviation, coefficient of variation, median, mode, average number of sentences, and Pearson's skew factor for each of the following: a) the Corpus; b) three grade corpora; c) seven subject corpora; d) eighteen subject within grade corpora; and e) thirty-seven textbook corpora. Comparative summary tables were developed for this data and are presented in Chapter IV.

Full details of the sentence length distribution of each of the sixty-six corpora are provided in a volume titled SENT. (See APPENDIX G). The volume arranges the data for each table under five headings. The first column (LENGTH) states the length of the sentence in words and the second column (REPETITIONS) gives the number of occurrences of this particular sentence length.

Column three (CUM. SENT) lists the sum of sentences counting from the top of the table and the fourth column (ACCUM WORDS) serves the same function for words. Column five (% WORDS) gives the running total of the percentage of tokens accounted for throughout the sentence length distribution.

Task 6.2 A matching set of graphs illustrating each of the sixty-six sentence length distributions was printed through the UBC plotting package using a CALCOMP Drum Plotter at the UBC Computing Centre. The graphs are presented as APPENDIX H.

Task 6.3 This task was designed to provide data on the repeat-rate frequency of sentence lengths for each of the following: a) the Corpus; b) the three grade corpora; seven subject corpora; eighteen subject within grade corpora; and e) thirty-seven textbook corpora.

The repeat-rate frequency tables for sentence lengths for each of the sixty-six corpora are also presented in SENT. A complete description of the sentence characteristics is available on tape #RE0616 at the UBC Computing Centre.

Task 6.3 made use of Yule's characteristic K along the lines suggested in Task 5.3 (Kucera and Francis, 1967). In this procedure, X in the statement $S1 = \sum_X fxX$, equals the number of occurrences of a specific sentence length and fx equals the number of cases of X. The characteristic K is useful in indicating whether material contains a great diversity of sentence lengths (low K value); or whether there is a high repetition of commonly occurring sentence lengths present (high K value). The implications of the K-factor for differences in writing style are discussed in later chapters. Summary tables of K for each of the sixty-six corpora outlined in Task 6.1 are presented in Chapter IV.

TASK 7. ANALYSIS OF DISTRIBUTION OF 100 MOST FREQUENT WORD-TYPES

Task 7 utilized existing computer programs and developed new programs where needed to analyze the distribution of the 100 most frequently occurring word-types across the following corpora: a) three grade levels; b) seven subject areas; c) six subjects in Grade 8; d) seven subjects in Grade 9; and e) five subjects in Grade 10. Two major sub-tasks were involved.

Task 7.1

The chi-square test was used to test whether there were significant differences in the distribution of the 100 most frequent word-types in the five areas described above, using the usual formula:

$$X^2 = \frac{(o - e)^2}{e}$$

where o = the observed frequency of the word-types, and e = the the expected frequency of the word-types. (The expected value equals the ratio of the total number of word-types in a corpora to the total number of word-types in the Corpus, multiplied by the total number of Corpus occurrences of the respective word-type being tested). The .01 level of significance was chosen to test the hypotheses in order to guard against a type 1 error. That is, it was decided to risk rejecting the null hypotheses when they were true only one time in a 100.

Complete details of the chi-square tests for the distribution of word-types have been arranged in a series of tables and are presented in APPENDIX I. For each word-type there are three lines of data. The first line gives the observed frequency, the second line lists the expected frequency, and the third line indicates the ratio of the number of occurrences of the specific word-type in the corpora to the total number of all word-types in the corpora expressed as a percentage. The 100 most frequent word-types are placed in ranked order on the left hand side of the tables.

Task 7.2 was designed to analyze and illustrate the number of word-types which differed significantly in their distribution across each of the five areas tested in Task 7.1. A summary table of these results is presented in Chapter IV. The table is organized into seven columns with the first column (RANK) giving the rank listing of each of the 100 word-types and the second column (WORD) listing the word-type. Columns three to seven indicate whether or not each of the word-types is evenly distributed across the three grade levels (GRADES); the seven subject areas (SUBJECTS C); the subjects in Grade 8 (SUBJECTS 8); the subjects in Grade 9 (SUBJECTS 9); and the subjects in Grade 10 (SUBJECTS 10).

TASK 8. ANALYSIS OF DISTRIBUTION OF SELECTED SENTENCE LENGTHS

This task involved the testing of a number of null hypotheses which stated that there were no significant differences in the sentence length distributions of the subject areas in the various corpora when compared to the normal population expressed by the sentence length distribution of the Corpus.

The chi-square test was used to test these hypotheses using the usual formula:

$$X^2 = \frac{(o - e)^2}{e}$$

where O= the observed frequency of the sentence length, and e= the expected frequency of the sentence length. (The expected value equals the ratio of the total number of sentence lengths in a corpora to the total number of sentence lengths in the Corpus, multiplied by the total number of Corpus occurrences of the respective sentence length being tested). The level of significance used to test these hypotheses was .01.

The chi-square tests were run using five ranges of sentence lengths : 10, 20, 30, 40, and 50+ words in length. These sentence lengths were chosen to represent short sentences, a group of sentences on either side of the Corpus mean sentence length, and two groups of longer sentences. The last range included all sentences with 50 words or above because of the variety and small number of sentences expected in this category. A computer program was developed to test the distribution of occurrence of the five selected sentence lengths across the three grade levels, the seven subject areas, the six subject areas in Grade 8, the seven subject areas in Grade 9, and the five subject areas of Grade 10. A summary of these results appears in Chapter IV. Complete details of the chi-square tests for the sentence length distribution have been arranged into five tables and are presented in APPENDIX J. The format of the tables is the same as that discussed in Task 7, except that the selected sentence lengths are placed on the left hand side of the tables.

TASK 9. IDENTIFICATION OF SIGNIFICANT CONTENT MATERIAL

The final task in the study involved the development of an "elimination technique" for the selection of the most significant words in specific content areas using the ranked frequency word lists developed for the Corpus, the three grade level corpora, the seven subject-area corpora, and the thirty-seven textbook corpora. Three sub-tasks were involved.

Task 9.1 Word frequency graphs were constructed for the eleven corpora described above using the UBC CALCOMP Drum Plotter. The graphs plotted the rank of each word-type along the abscissa and the frequency of each word-type on the ordinate. Because of the magnitude of the quantities being plotted, a one-tenth scale was used. (See APPENDIX K). The word frequency graphs take the general shape of the diagram in FIGURE 3.

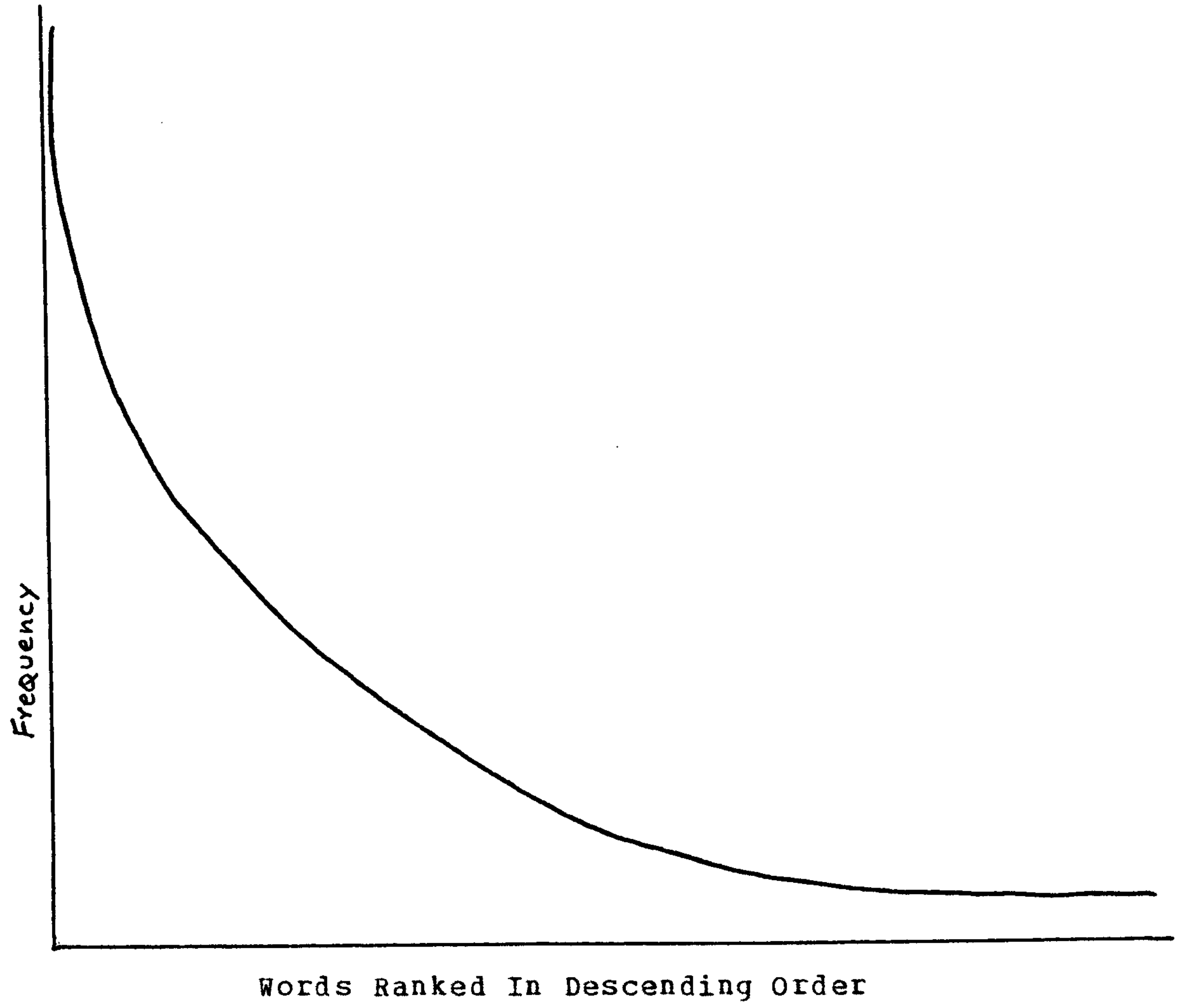


FIGURE 3

MODEL OF A WORD FREQUENCY DIAGRAM

Task 9.2 The "elimination technique" suggested in this study consists of two stages. The first stage is designed to identify the high frequency words in a word list that are considered to be too common to have special significance for the content area under investigation. A cutoff point is determined and these high frequency words are eliminated. The decision was made to use the position on the abscissa where 50 percent of the tokens occurred as the cutoff point. This involves elimination of most of the structure words. The line A in FIGURE 4 represents this cutoff.

The second stage is designed to eliminate words which are too rare or do not have sufficient frequency of occurrence to warrant their being considered as significant for the specific content area. A cutoff point is determined and these low frequency words are eliminated. The decision was made to use the position on the abscissa where approximately 10 percent of the low frequency tokens occurred as the cutoff point. This results in the elimination of words that occur only one to three times in most lists and which are regarded to be low in significance. The line B in FIGURE 4 represents the cutoff. Words which fall in the gray area immediately to the left and right of both point A and B could of course also be included as having significance depending on the judgment of the individual selecting significant content and the degree of accuracy desired in designating the words to be eliminated.

Task 9.3 The balance of the words remaining between points A and B (approximately 40 percent of the total tokens), represents, for most purposes, the most significant content in a word list. That is, these are the items of vocabulary that occur neither too frequently to be classed as common words, nor too infrequently to be classed as rare words. It must again be emphasized that subjective judgment by specialists in the content area concerned is vital in making final decisions in the elimination and retention of 'gray' area words and in establishing the general cutoff points for A and B.

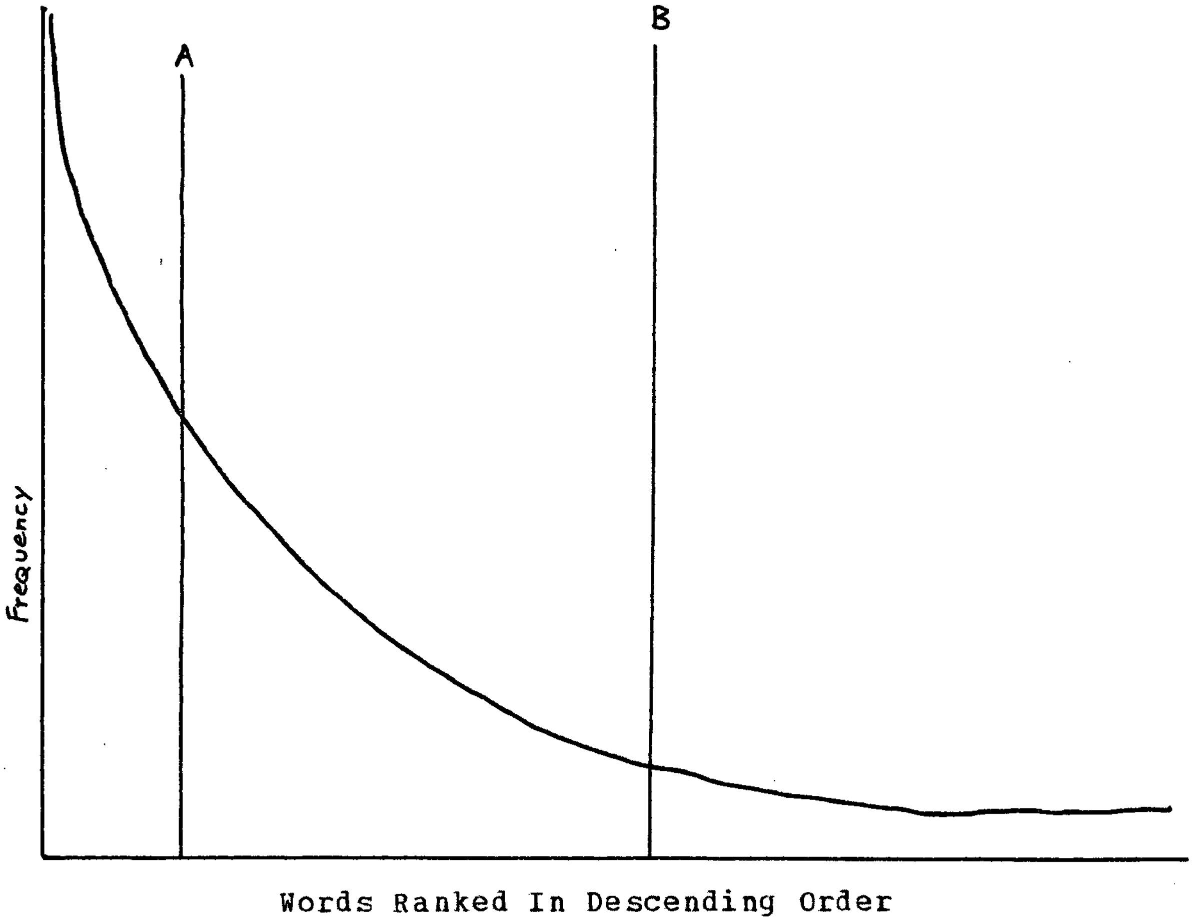


FIGURE 4

APPLICATION OF "ELIMINATION TECHNIQUE" TO THE MODEL OF A WORD FREQUENCY DIAGRAM

A complete discussion of the results of applying the "elimination technique" to the Corpus, the three grade level corpora, and the seven subject-area corpora is presented in Chapter IV using the frequency distribution graph of the Corpus as an example.

One final point should be made with respect to the design and methodology of the study. The entire production of the dissertation was accomplished through the use of computer technology. Initially, each chapter was keypunched onto IBM computer cards, read into the computer memory bank and stored on disk. The dissertation was then edited, using a 3270 CRT unit, revised numerous times and finally printed in its present form using the FMT computer program. The graphs and chi-square tables in the APPENDIX were produced by special programs and reduced for convenience of presentation.

The use of the computer in producing the dissertation had the great advantage of allowing constant revisions to be made and multiple copies of the revised manuscript to be obtained very quickly. The formatting of tables and other descriptive statistics plus the construction of graphs were also relatively easy with computer facilities. The major drawback in using computer techniques was the need for the researcher to edit very carefully the 'logical' but set procedures used by the computer in the organization and interpretation of print materials. This involved an understanding of basic computer processes plus some of the programming language used in generating the computer output.

CHAPTER IV

ANALYSIS OF THE DATA AND FINDINGS

The purpose of the chapter is to present and analyze the results obtained from the completion of Tasks 1-9. The tasks resulted in the production of over 5,500 pages of printed material, including print facsimiles of all the instructional materials sampled, the sixty-six word lists, and accompanying tables, graphs, and statistical summaries. These data were then organized into eight volumes and are discussed fully in Tasks 3 and 4.

All of the material generated in the study, including over 200 computer files used to organize the material and twenty specially developed computer programs, have been placed on magnetic tape. Copies of the tape are available from the Computing Centre (Tape #RE0616) and the Special Collections Division of the Library (Tape #RE0617) at the University of British Columbia. A technical description of the procedures followed in developing and using the various computer programs is given in the booklet, Programmer's Guide to the Edwards' Corpus, by Allan Miller, available from the Computing Centre at the University of British Columbia.

TASKS, QUESTIONS AND HYPOTHESES

The tasks outlined in Chapter I are restated in this section, the questions answered, the hypotheses tested, and the general findings presented.

Task 1.

Develop a representative corpus of natural language text based on instructional material prescribed for use in British Columbia junior secondary grades.

The thirty-seven textbooks and 469 samples used in developing the 235,107 word Corpus of instructional materials are described in APPENDIX A. The sample sizes ranged from 657 words to 338 words with a mean of 501.294 and a standard deviation of 44.187. Two copies of the 469 sample sizes, one organized in alphabetical order, and one ranked by size in ascending order are presented in APPENDIX B.

Task 2.

Organize the Corpus for computer input and manipulation.

The keypunching of computer cards containing the Corpus was accomplished using the UBC FMT (FORMAT) program available from the Computing Centre at the University of British Columbia. The computer cards were read into the computer via a card-reader and placed on disk to await reorganization into the various tasks involved in the study.

Task 3.

Generate two volumes of the Corpus: one organized by grade levels, and one organized by subject areas, each with a descriptive index.

Computer programs were utilized to generate the two volumes of the Corpus: 1) C.G., which presents print facsimiles of the instructional material organized by grade levels, and 2) C.S., which organizes the instructional material by subject. A description of the development of the corpora, which includes an index and full particulars for each text and the samples used, is included in the front of each of the two volumes. A detailed listing of the 209 computer files and programs used in the study is presented in APPENDIX C.

Task 4.

Organize the samples into word lists for the Corpus, the grade corpora (3), the subject corpora (7), the subjects within grade corpora (18), and the textbook corpora (37).

4.1 For each of the above, provide an alphabetical and a rank order (descending frequency) listing of word-types to give the following information.

4.11 The frequency of occurrence of each word-type.
4.12 The cumulative percentage frequency of each word-type.
4.13 The relative frequency of occurrence of each word-type per 1000 tokens.
4.14 The descriptive statistics for the rank order lists of the Corpus and corpora including: X, FX, SUM FX, FX * X, CUM % FX * X .

4.2 Construct two summary tables for each of the sixty-six word lists, indicating the word frequency figures in descending order (highest frequency words first), and in ascending order (highest frequency words last).

Task 4.1 The alphabetical and rank order word lists and relevant statistical details for the Corpus and all corpora are organized into five volumes as follows:

1) C.V. Represents the word list for the Corpus (345

pages),

2) G.V. Represents the word lists for the three grade level corpora (550 pages),

3) S.V. Represents the word lists for the seven subject area corpora (730 pages),

4) S.G.V. Represents the word lists for the eighteen subject within grade level corpora (986 pages),

5) T.V. Represents the word lists for the thirty-seven textbook corpora (1,292 pages).

All word lists are set up in two columns per page with fifty words per column for added convenience. The organization of the alphabetical lists and the rank order lists is basically the same for all corpora with one exception. Each word entry in both lists is preceded by two figures. For the alphabetical list the first quantity in column FREQ indicates the relative frequency per 1000 tokens of the word entry. For the rank order list the first quantity in the FREQ column indicates the cumulative total of the tokens in the Corpus contributed by the word entry. The second figure in each list gives the frequency of the word entry. (See APPENDIXES D and E).

The alphabetical list begins with several command symbols plus a complete listing of the alphanumerical indexes of the 469 samples used in the study. All other types that do not begin with letters are placed at the end of the list.

The rank order list begins with the highest frequency word in the Corpus and places all other types in descending rank. <u>Where more than one type has the same frequency of occurrence</u>

<u>the order within the respective frequency is alphabetical with numbers listed last.</u> A complete listing of the alphanumerical indexes of the 469 samples used in the study appears at the beginning of the <u>hapax legomena</u> entries.

<u>Task 4.2</u> The descending word frequency lists constructed for each of the sixty-six corpora gives the rank of each word-type in descending order. A sample page from the descending lists is included in APPENDIX F. This list enables the rank of any word to be quickly located by first finding the frequency of occurrence of the word in the alphabetical list and then matching this number with the same frequency in column X of the descending list. For example, if the rank of the word "about" in the Corpus is required, the reader could note that the frequency of the word in the alphabetical list (APPENDIX D) is 463 and determine from "The Corpus with Rank in Descending Order" list (APPENDIX F), that X = 463 corresponds to a rank of 55 which is the rank of the word "about" in the Corpus. This means that there are 54 words which have a greater frequency of occurrence than the word "about" and 16,350 words which occur less frequently in the Corpus. A similar procedure could be followed with entries in any of the other sixty-five corpora word lists.

Another service offered by the descending order word frequency list involves determining the relationship between word-types and tokens. The descending list for the Corpus indicates that the first 100 most frequent words constitute only 0.610 percent of the word-types in the Corpus yet the same words account for 115,141 tokens or 48.973 percent of the total number of word occurrences in the Corpus.

The ascending order word frequency list developed for each corpora gives the rank of each word-type in ascending order. A sample page is included in APPENDIX F. This list is useful in determining the number of tokens accounted for by the rarely occurring word-types. For example, "The Corpus in Ascending Order" list, indicates that low frequency word-types occurring ten times or less constitute 84.505 percent of the word-types in the Corpus yet account for only 34,572 tokens or 14.705 percent of the total number of word occurrences in the Corpus.

Task 5

Generate comparative and statistical analyses based on the lexical characteristics of the Corpus, the corpora, and data produced in Tasks 1 through 4.

5.1 What are the lexical characteristics of the Corpus; the Grade 8,9, and 10 corpora; each of the seven subject area corpora across Grades 8, 9, and 10; each of the subject corpora within Grade 8, 9, and 10; and each of the thirty-seven textbook corpora, in terms of the total number of graphic characters, average number of graphic characters per token, tokens and discrete word-types?

5.2 What are the characteristics in terms of repeat-rate frequency (Yule's K) of words for the Corpus and corpora defined in Task 5.1?

Task 5.1 The lexical characteristics of the Corpus and the various corpora are presented in TABLES IV through X.

The total Corpus includes 16,405 word-types across the 469 samples developed for the study. TABLE IV illustrates the relatively large size of the Grade 9 corpus in contrast to those of Grades 8 and 10. Over 50 percent (122,953) of the tokens in the total Corpus are represented by 69 percent (11,401) of the

Corpus word-types in the nineteen textbooks used in Grade 9. The Grade 8 (52,867 tokens) and Grade 10 (59,343 tokens) corpora are approximately the same size in terms of both word-types and tokens.

TABLE IV

NUMBER OF TYPES, TOKENS, CHARACTERS, AND AVERAGE LENGTH OF TOKENS IN CHARACTERS FOR GRADE LEVELS AND THE CORPUS

Grade	Types	Tokens	Characters	Average
8	7,027	52,867	234,527	4.44
9	11,401	122,953	554,488	4.51
10	7,736	59,343	273,654	4.61
Corpus	16,405	235,107	1,062,411	4.52

The lexical characteristics of the subject areas of the Corpus across the three grade levels, outlined in TABLE V, indicate that Home Economics (49,257 tokens) is the largest subject corpora and Mathematics (17,808) the smallest. English, which is the second largest corpora (40,300 tokens) has by far the most word-types (7,079) indicating a much greater variety of vocabulary used throughout the eight textbooks in this subject as compared to the other content areas in the junior secondary grades.

TABLE V

NUMBER OF TYPES, TOKENS, CHARACTERS, AND AVERAGE LENGTH OF TOKENS IN CHARACTERS FOR THE SUBJECT AREAS ACROSS GRADE LEVELS

Subject	Types	Tokens	Characters	Average
Commerce	3,020	20,155	90,171	4.47
English	7,079	40,300	178,192	4.42
Home Economics	5,529	49,257	221,576	4.50
Industrial Ed.	4,060	31,300	141,176	4.51
Mathematics	1,952	17,808	73,852	4.15
Science	4,833	37,787	173,023	4.58
Social Studies	6,211	38,608	184,727	4.78
Corpus	16,405	235,107	1,062,411	4.52

TABLE VI gives the lexical characteristics of the six subject areas (Commerce is not offered) within Grade 8. Home Economics and Social Studies are the two largest corpora with over 11,000 tokens each although English has a greater number of word-types (2,388) than Home Economics (2,169). Social Studies, although ranking second in total tokens, has the greatest number of word-types for Grade 8 (2,890 types).

TABLE VI

NUMBER OF TYPES, TOKENS, CHARACTERS, AND AVERAGE LENGTH OF TOKENS IN CHARACTERS FOR THE SUBJECT AREAS OF GRADE 8

Subject	Types	Tokens	Characters	Average
Commerce	-	-	-	-
English	2,388	8,605	37,901	4.40
Home Economics	2,169	11,425	50,472	4.42
Industrial Ed.	1,305	4,624	20,981	4.54
Mathematics	1,164	7,073	30,201	4.27
Science	1,975	9,907	43,363	4.38
Social Studies	2,890	11,205	51,480	4.59
Grade 8	7,027	52,867	234,527	4.44

In Grade 9, (TABLE VII), Home Economics is the largest corpus (37,812 tokens) followed by Industrial Education (26,656 tokens) and English (23,123 tokens). English again has the largest number of word-types. Only one Mathematics text was used (the algebra text was excluded) resulting in a relatively small number of samples of running prose (3,616 tokens). Grade 9 is the largest of the three grade level corpora with a total of nineteen textbooks included.

TABLE VII

NUMBER OF TYPES, TOKENS, CHARACTERS, AND AVERAGE LENGTH OF TOKENS IN CHARACTERS FOR THE SUBJECT AREAS OF GRADE 9

Subject	Types	Tokens	Characters	Average
Commerce	2,208	12,485	55,653	4.46
English	4,920	23,123	103,490	4.48
Home Economics	4,894	37,812	171,040	4.52
Industrial Ed.	3,688	26,656	120,125	4.51
Mathematics	910	3,616	15,460	4.28
Science	2,365	12,278	55,612	4.53
Social Studies	2,065	6,955	32,973	4.74
Grade 9	11,401	122,953	273,654	4.51

TABLE VIII lists the lexical characteristics of the five subjects in Grade 10. The largest subject corpus in Grade 10 is Social Studies (20,428 tokens) and the smallest is Mathematics (7,100 tokens). Social Studies also has the largest number of word-types (3,930).

The Grade 9 textbooks for Home Economics and Industrial Education are repeated in Grade 10 but were not used again in the study. Nine textbooks were used to obtain samples and six textbooks were excluded because they did not contain sufficiently large quantities of running prose. The six excluded texts included two Commerce books, two English books, a Mathematics (Geometry) text, and an atlas used in Social Studies.

TABLE VIII

NUMBER OF TYPES, TOKENS, CHARACTERS, AND AVERAGE LENGTH OF TOKENS IN CHARACTERS FOR THE SUBJECT AREAS OF GRADE 10

Subject	Types	Tokens	Characters	Average
Commerce	1,746	7,651	34,459	4.50
English	2,489	8,553	36,743	4.30
Home Economics	-	-	-	-
Industrial Ed.	-	-	-	-
Mathematics	912	7,100	28,123	3.96
Science	3,015	15,583	73,990	4.75
Social Studies	3,930	20,428	100,210	4.91
Grade 10	7,736	59,343	273,654	4.61

A summary of all the lexical characteristics for each of the subject areas across grade levels is presented in TABLE IX . The largest selection of material at the one grade level dealt with in this study was in Grade 9 Home Economics (37,812 tokens) where five textbooks were sampled. Other subject areas containing large amounts of running prose were Grade 9 Industrial Education (26,656 tokens), Grade 9 English (23,123 tokens), and Grade 10 Social Studies (20,428 tokens). The smallest quantities of running prose were located in Grade 9 Mathematics (3,616 tokens), Grade 8 Industrial Education (4,624 tokens), and Grade 9 Social Studies (6,955 tokens). Grade 10 Mathematics contained the smallest recorded 'average length of tokens in characters' (3.96) throughout the study. The largest number of word-types occur in Grade 9 English (4,920) where four textbooks were sampled. Grade 9 Home Economics is a close second with 4,894 types.

TABLE IX

NUMBER OF TYPES, TOKENS, CHARACTERS, AND AVERAGE LENGTH OF TOKENS IN CHARACTERS FOR THE SUBJECT AREAS OF EACH GRADE LEVEL OF THE CORPUS

Subject	Grade	Types	Tokens	Characters	Average
Commerce	9	2,208	12,485	55,653	4.46
Commerce	10	1,746	7,651	34,459	4.50
English	8	2,388	8,605	37,901	4.40
English	9	4,920	23,123	103,490	4.48
English	10	2,489	8,553	36,743	4.30
Home Economics	8	2,169	11,425	50,472	4.42
Home Economics	9	4,894	37,812	171,040	4.52
Industrial Ed.	8	1,305	4,624	20,981	4.54
Industrial Ed.	9	3,688	26,656	120,125	4.51
Mathematics	8	1,164	7,073	30,201	4.27
Mathematics	9	910	3,616	15,460	4.28
Mathematics	10	912	7,100	28,123	3.96
Science	8	1,975	9,907	43,363	4.38
Science	9	2,365	12,278	55,612	4.53
Science	10	3,015	15,583	73,990	4.75
Soc. Studies	8	2,890	11,205	51,480	4.59
Soc. Studies	9	2,065	6,955	32,973	4.74
Soc. Studies	10	3,930	20,428	100,210	4.91

TABLE X lists the lexical characteristics of each of the thirty-seven textbooks used in the study. A Grade 10 Social Studies text (*3H01), _A Regional Geography of North America,_ contains the largest selection of running prose (14,736 tokens), while _Drama IV,_ a Grade 10 English text (*3C02), has the smallest number of tokens (1,867). A Grade 10 Social Studies textbook has the largest number of word-types (2,913) and a Grade 10 English test has the least (822).

TABLE X

NUMBER OF TYPES, TOKENS, CHARACTERS, AND AVERAGE LENGTH OF TOKENS IN CHARACTERS FOR THE THIRTY-SEVEN TEXTS

Text	Types	Tokens	Characters	Average
*1C01 (Eng)	1,187	3,500	15,601	4.46
*1C02 (Eng)	1,672	5,105	22,300	4.37
*1D01 (H.Ec)	2,169	11,425	5,8472	4.42
*1E01 (I.Ed)	1,305	4,624	20,981	4.54
*1F01 (Math)	1,164	7,073	30,201	4.27
*1G01 (Sci)	1,033	4,402	18,926	4.30
*1G02 (Sci)	1,399	5,505	24,437	4.44
*1H01 (S.St)	2,177	7,728	35,235	4.56
*1H02 (S.St)	1,215	3,477	16,245	4.67
*2B01 (Comm)	1,234	5,494	24,022	4.37
*2B02 (Comm)	1,511	6,991	31,631	4.52
*2C01 (Eng)	2,436	9,646	44,736	4.64
*2C02 (Eng)	1,232	3,400	15,122	4.45
*2C03 (Eng)	1,705	5,035	21,839	4.34
*2C04 (Eng)	1,638	10,198	21,793	4.32
*2D01 (H.Ec)	1,872	10,755	46,425	4.55
*2D02 (H.Ec)	1,871	6,928	48,323	4.49
*2D03 (H.Ec)	1,685	4,599	31,352	4.53
*2D04 (H.Ec)	1,467	5,332	24,051	4.51
*2D05 (H.Ec)	1,269	4,599	20,889	4.54
*2E01 (I.Ed)	1,615	6,075	27,547	4.53
*2E02 (I.Ed)	1,638	7,792	34,579	4.44
*2E03 (I.Ed)	2,062	12,789	57,999	4.54
*2F01 (Math)	910	3,616	15,466	4.28
*2G01 (Sci)	1,516	6,748	30,618	4.54
*2G02 (Sci)	1,474	5,530	24,994	4.52
*2H01 (S.St)	1,420	4,408	20,365	4.62
*2H02 (S.St)	984	2,547	12,608	4.95
*3B01 (Comm)	1,017	3,546	15,477	4.36
*3B02 (Comm)	1,170	4,105	18,982	4.62
*3C01 (Eng)	1,946	6,686	27,972	4.18
*3C02 (Eng)	822	1,867	8,771	4.70
*3F01 (Math)	912	7,100	28,123	3.96
*3G01 (Sci)	1,955	8,592	40,616	4.73
*3G02 (Sci)	1,844	6,991	33,374	4.77
*3H01 (S.St)	2,913	14,736	70,766	4.80
*3H02 (S.St)	1,837	5,692	29,444	5.17

Task 5.2 The repeat-rate frequency tables for the Corpus and corpora are listed in the five volumes C.V., G.V., S.V., S.G.V., and T.V. (See Task 4.2). The results for Yule's characteristic K (for words) are presented in TABLES XI through XVI.

As stated earlier in Chapter III, the K value is useful as a measure of the repeat rate of words and provides an indication of the concentration of vocabulary in a passage of printed material. A large K factor suggests a proportionately greater use of high frequency (common) words than a small value of K which implies more reliance on low frequency (rare) words.

The K factor is theoretically independent of sample size when the samples have been randomly selected from the population being used. For this reason it is possible to compare the results of K for the various corpora.

The K factors for each grade level and the Corpus are presented in TABLE XI. Grade 9 has the smallest value of K (106.547) and Grade 10 has the largest K value (112.587) although all grades were close to the K value for the Corpus (108.104).

TABLE XI

K FACTORS (WORDS) FOR EACH GRADE LEVEL AND THE CORPUS

Grade	K Factor
8	109.510
9	106.547
10	112.587
Corpus	108.104

TABLE XII presents the K factors ranked for the subject areas across grade levels. Home Economics (92.572) and English (100.517) have markedly lower values of K implying that these subjects use a relatively greater number of low frequency (rare) words than the other subjects.

TABLE XII

SUBJECT AREAS ACROSS GRADES RANKED BY K FACTOR (WORDS)

Rank	Subject	K Factor
1	Home Economics	92.572
2	English	100.517
3	Corpus	108.104
4	Commerce	108.922
5	Mathematics	121.662
6	Science	129.894
7	Industrial Education	129.922
8	Social Studies	130.372

The K factors for the subject areas within each of the three grade levels are presented in TABLE XIII and their ranked order is listed in TABLE XIV. In English, Mathematics, and Social Studies, the lowest value of K occurs in Grade 9, indicating a greater use of low frequency (rare) words than in either of the other two grades, while in Industrial Education and Science, the lowest K values are in Grade 8 and Grade 10 respectively. In Home Economics and Commerce, the lowest K values are in Grades 9 and 10 respectively. Four out of the seven subjects have their lowest K values in Grade 9 with two in Grade 10 and one in Grade 8.

TABLE XIII

K FACTORS (WORDS) FOR SUBJECTS WITHIN GRADE LEVELS

Subject	Gde 8	Gde 9	Gde 10	Corpus
Commerce	-	117.619	99.329	108.922
English	107.175	98.491	104.271	100.517
Home Economics	98.166	91.788	-	92.572
Industrial Ed.	116.973	133.630	-	129.922
Mathematics	123.568	118.672	131.571	121.662
Science	135.992	145.159	118.004	129.894
Social Studies	130.613	127.738	133.350	130.372
Corpus	109.510	106.547	112.587	-

TABLE XIV presents the rank of the subject areas within grades and indicates that Commerce, Home Economics, and English occupy seven of the first eight places among the eighteen positions.

TABLE XIV

SUBJECT AREAS WITHIN GRADE LEVELS RANKED BY K FACTOR (WORDS)

Rank	Subject	Gde	K Factor
1.	Home Economics	9	91.788
2.	Home Economics	8	98.166
3.	English	9	98.491
4.	Commerce	10	99.329
5.	English	10	104.271
6.	English	8	107.175
7.	Industrial Education	8	116.973
8.	Commerce	9	117.619
9.	Science	10	118.004
10.	Mathematics	9	118.672
11.	Mathematics	8	123.568
12.	Social Studies	9	127.738
13.	Social Studies	8	130.613
14.	Mathematics	10	131.571
15.	Social Studies	10	133.350
16.	Industrial Education	9	133.630
17.	Science	8	135.992
18.	Science	9	145.159

The K factors for each individual textbook follow. They are: ranked by subject areas (TABLE XV), listed by subjects within a grade level (TABLE XVI), and ranked independently across all subjects and grade levels (TABLE XVII).

The low K values for the textbooks in Home Economics and English is evident in TABLE XV. Only one of the Home Economics texts, <u>Guide to Modern Meals</u> (*2D01), has a K factor over 100, while most of the English texts have K factors approaching 100.

TABLE XV

TEXTS IN SUBJECT AREAS RANKED BY K FACTOR (WORDS)

Text	Subject	K Factor
*3B01	Commerce	95.532
*3B02	Commerce	111.939
*2B01	Commerce	114.560
*2B02	Commerce	125.988
*2C03	English	99.231
*2C04	English	99.253
*1C02	English	101.873
*3C01	English	102.632
*2C02	English	103.655
*2C01	English	105.651
*1C01	English	118.117
*3C02	English	125.065
*2D04	Home Economics	81.857
*2D03	Home Economics	87.723
*2D05	Home Economics	92.203
*2D02	Home Economics	97.723
*1D01	Home Economics	98.166
*2D01	Home Economics	111.747
*2E01	Industrial Education	113.084
*2E02	Industrial Education	114.428
*1E01	Industrial Education	116.973
*2E03	Industrial Education	169.462
*2F01	Mathematics	118.672
*1F01	Mathematics	123.568
*3F01	Mathematics	131.571
*1G02	Science	117.664
*3G02	Science	117.712
*3G01	Science	128.905
*2G02	Science	142.048
*2G01	Science	150.283
*1G01	Science	167.198
*2H01	Social Studies	126.723
*1H02	Social Studies	127.347
*3H02	Social Studies	128.258
*2H02	Social Studies	134.655
*1H01	Social Studies	137.026
*3H01	Social Studies	137.962

TABLE XVI

K FACTOR (WORDS) FOR EACH TEXT BY GRADES

Text	Subject	K Factor
*1C01	English	118.117
*1C02	English	101.873
*1D01	Home Economics	98.166
*1E01	Industrial Education	116.973
*1F01	Mathematics	123.568
*1G01	Science	167.198
*1G02	Science	117.664
*1H01	Social Studies	137.026
*1H02	Social Studies	127.347
*2B01	Commerce	114.560
*2B02	Commerce	125.988
*2C01	English	105.651
*2C02	English	103.655
*2C03	English	99.231
*2C04	English	99.253
*2D01	Home Economics	111.747
*2D02	Home Economics	97.723
*2D03	Home Economics	87.723
*2D04	Home Economics	81.857
*2D05	Home Economics	92.203
*2E01	Industrial Education	113.084
*2E02	Industrial Education	114.428
*2E03	Industrial Education	169.462
*2F01	Mathematics	118.672
*2G01	Science	150.283
*2G02	Science	142.048
*2H01	Social Studies	126.723
*2H02	Social Studies	134.655
*3B01	Commerce	95.532
*3B02	Commerce	111.939
*3C01	English	102.632
*3C02	English	125.065
*3F01	Mathematics	131.571
*3G01	Science	128.905
*3G02	Science	117.712
*3H01	Social Studies	137.962
*3H02	Social Studies	128.258

Within Grade 8 and 9, Home Economics has the lowest K value while Commerce has the lowest value within Grade 10.

TABLE XVII presents the ranked order of the textbooks by K factor (words). Five of the first twelve textbooks are Home Economics, six are English texts, and one is a Commerce text.

TABLE XVII

K FACTOR (WORDS) FOR EACH TEXT RANKED ACROSS SUBJECTS AND GRADES

Rank	Text	Subject	K Factor
1.	*2D04	Home Economics	81.857
2.	*2D03	Home Economics	87.723
3.	*2D05	Home Economics	92.203
4.	*3B01	Commerce	95.532
5.	*2D02	Home Economics	97.723
6.	*1D01	Home Economics	98.166
7.	*2C03	English	99.231
8.	*2C04	English	99.253
9.	*1C02	English	101.873
10.	*3C01	English	102.632
11.	*2C02	English	103.655
12.	*2C01	English	105.651
13.	*2D01	Home Economics	111.747
14.	*3B02	Commerce	111.939
15.	*2E01	Industrial Education	113.084
16.	*2E02	Industrial Education	114.428
17.	*2B01	Commerce	114.560
18.	*1E01	Industrial Education	116.973
19.	*1G02	Science	117.664
20.	*3G02	Science	117.712
21.	*1C01	English	118.117
22.	*2F01	Mathematics	118.672
23.	*1F01	Mathematics	123.568
24.	*3C02	English	125.065
25.	*2B02	Commerce	125.988
26.	*2H01	Social Studies	126.723
27.	*1H02	Social Studies	127.347
28.	*3H02	Social Studies	128.258
29.	*3G01	Science	128.905
30.	*3F01	Mathematics	131.571
31.	*2H02	Social Studies	134.655
32.	*1H01	Social Studies	137.026
33.	*3H01	Social Studies	137.962
34.	*2G02	Science	142.048
35.	*2G01	Science	150.283
36.	*1G01	Science	167.198
37.	*2E03	Industrial Education	169.462

Task 6.

Generate comparative and statistical analyses based on the sentence length distribution of the Corpus, the corpora, and data produced in Task 1 through Task 4.

6.1 What are the sentence-length characteristics of the Corpus; the Grade 8, 9, and 10 corpora; each of the seven subject area corpora across Grades 8, 9, and 10; each of the corpora for subjects within Grades 8, 9, and 10; and each of the thirty-seven textbook corpora in terms of the mean, median, modal sentence length in words, standard deviation, coefficient of variation, average number of sentences, and Pearson's skew factor.

6.2 Produce a set of graphs to illustrate each of the sixty-six sentence length distributions developed during the study.

6.3 What are the characteristics in terms of repeat-rate frequency of sentence lengths (Yule's K) for the Corpus and the corpora defined in 6.1 above?

Task 6.1 The sentence length characteristics of the Corpus and the various corpora (all measured in number of words) are given in TABLES XVIII through XXIV. Complete details of the sentence-length distribution of the Corpus and each of the sixty-six corpora are presented in the volume SENT. A sample of the contents of SENT is included in APPENDIX G.

TABLE XVIII illustrates the fairly uniform average sentence length when the samples are organized by grade levels across the Corpus. This pattern is also repeated when the samples are organized by subjects across the three grades (TABLE XIX), although the range in averages increases.

TABLE XVIII

MEAN SENTENCE LENGTH, STANDARD DEVIATION, COEFFICIENT OF VARIATION, MEDIAN, MODE, AND AVERAGE NUMBER OF SENTENCES PER SAMPLE FOR EACH GRADE LEVEL AND THE CORPUS

Grade	Mean	S.D.	Variation	Median	Mode	Average Sentences
8	18.595	9.7745	0.5256	16.764	18	27.33
9	17.824	10.2550	0.5753	15.428	15	28.04
10	17.593	9.8504	0.5599	15.733	10	28.34
Corpus	17.927	10.0480	0.5605	15.743	15	27.96

TABLE XIX

MEAN SENTENCE LENGTH, STANDARD DEVIATION, COEFFICIENT OF VARIATION, MEDIAN, MODE, AND AVERAGE NUMBER OF SENTENCES PER SAMPLE FOR EACH SUBJECT AREA ACROSS GRADES

Subject	Mean	S.D.	Variation	Median	Mode	Average Sentences
Commerce	17.772	9.080	0.510	15.770	13	27.66
English	17.568	13.685	0.779	13.750	7	28.68
Home Economics	18.476	8.633	0.467	16.813	16	27.20
Industrial ed.	16.683	8.449	0.506	14.550	11	29.78
Mathematics	15.247	8.150	0.534	13.532	14	33.37
Science	18.495	9.785	0.529	16.444	15	27.24
Social studies	19.973	9.582	0.479	18.207	21	25.10
Corpus	17.927	10.048	0.560	15.743	15	27.96

TABLE XX presents the sentence length characteristics across the grade levels. The smallest average sentence length is in Grade 8 Mathematics and the largest in Grade 10 Social Studies.

TABLE XX

MEAN SENTENCE LENGTH, STANDARD DEVIATION, COEFFICIENT OF VARIATION, MEDIAN, MODE, AND AVERAGE NUMBER OF SENTENCES PER SAMPLE FOR EACH SUBJECT AREA WITHIN GRADE LEVELS OF THE CORPUS

Subject	Gde	Mean	S.D.	Var.	Median	Mode	Average Sentences
Comm.	9	17.558	9.642	0.549	15.475	17	28.440
Comm.	10	18.087	8.016	0.443	16.300	14	26.437
Eng.	8	17.597	13.049	0.741	14.280	12	28.764
Eng.	9	18.753	14.585	0.777	14.509	8	26.234
Eng.	10	14.953	11.670	0.780	11.230	7	35.750
H.Ec.	8	19.430	8.105	0.417	18.100	16	26.727
H.Ec.	9	18.196	8.738	0.480	16.442	16	27.342
I.Ed.	8	17.511	7.677	0.438	15.530	14	29.333
I.Ed.	9	16.535	8.552	0.517	14.392	15	29.851
Math.	8	17.421	8.872	0.509	16.170	18	29.000
Math.	9	14.406	6.802	0.472	13.190	9	35.857
Math.	10	13.894	7.781	0.560	12.330	10	36.500
Sci.	8	17.081	8.631	0.505	15.170	15	29.000
Sci.	9	17.924	9.173	0.511	15.757	15	28.541
Sci.	10	20.028	10.861	0.542	17.900	18	25.096
S.St.	8	21.715	9.876	0.454	19.700	23	23.454
S.St.	9	21.204	11.137	0.525	18.444	15	25.230
S.St.	10	18.758	8.666	0.462	17.260	10	28.340

The sentence length characteristics for subjects within Grade 8, Grade 9, and Grade 10 are presented in TABLES XXI, XXII, and XXIII respectively.

TABLE XXI

MEAN SENTENCE LENGTH, STANDARD DEVIATION, COEFFICIENT OF VARIATION, MEDIAN, MODE, AND AVERAGE NUMBER OF SENTENCES PER SAMPLE FOR THE SUBJECT AREAS OF GRADE 8

Subject	Mean	S.D.	Variation	Median	Mode	Average Sentences
Commerce						
English	17.597	13.049	0.741	14.280	12	28.764
Home Economics	19.430	8.105	0.417	18.100	16	26.727
Industrial Ed.	17.511	7.677	0.438	15.530	14	29.333
Mathematics	17.421	8.872	0.509	16.170	18	29.000
Science	17.081	8.631	0.505	15.170	15	29.000
Social Studies	21.715	9.876	0.454	19.700	23	23.454
Grade 8	18.595	9.774	0.525	16.764	18	27.330

TABLE XXII

MEAN SENTENCE LENGTH, STANDARD DEVIATION, COEFFICIENT OF VARIATION, MEDIAN, MODE, AND AVERAGE NUMBER OF SENTENCES PER SAMPLE FOR THE SUBJECT AREAS OF GRADE 9

Subject	Mean	S.D.	Variation	Median	Mode	Average Sentences
Commerce	17.558	9.642	0.549	15.475	17	28.440
English	18.753	14.585	0.777	14.509	8	26.234
Home Economics	18.196	8.738	0.480	16.442	16	27.342
Industrial Ed.	16.535	8.552	0.517	14.392	15	29.851
Mathematics	14.406	6.802	0.472	13.190	9	35.857
Science	17.924	9.173	0.511	15.757	15	28.541
Social Studies	21.204	11.137	0.525	18.444	15	25.230
Grade 9	17.824	10.255	0.575	15.428	15	28.040

TABLE XXIII

MEAN SENTENCE LENGTH, STANDARD DEVIATION, COEFFICIENT OF VARIATION, MEDIAN, MODE, AND AVERAGE NUMBER OF SENTENCES PER SAMPLE FOR THE SUBJECT AREAS OF GRADE 10

Subject	Mean	S.D.	Variation	Median	Mode	Average Sentences
Commerce	18.087	8.016	0.443	16.30	14	26.437
English	14.953	11.670	0.780	11.23	7	35.750
Home Economics	-	-	-	-	-	-
Industrial Ed.	-	-	-	-	-	-
Mathematics	13.894	7.781	0.560	12.33	10	36.500
Science	20.028	10.861	0.542	17.90	18	25.096
Social Studies	18.758	8.666	0.462	17.26	21	25.928
Grade 10	17.593	9.850	0.559		10	28.340

The average sentence lengths for the subject areas within grades differ considerably with Grade 9 exhibiting the greatest range in sentence lengths.

TABLE XXIV lists the sentence length characteristics for the thirty- seven textbooks.

TABLE XXIV

MEAN SENTENCE LENGTH, STANDARD DEVIATION, COEFFICIENT OF VARIATION, MEDIAN, MODE, AND AVERAGE NUMBER OF SENTENCES PER SAMPLE FOR THE THIRTY-SEVEN TEXTS

Subject	Mean	S.D.	Variation	Median	Mode	Average Sentences
*1C01 (Eng.)	18.717	14.639	0.782	14.60	15	26.714
*1C02 (Eng.)	16.904	11.933	0.706	14.00	12	30.200
*1D01 (H.Ec)	19.430	8.105	0.417	18.10	16	26.727
*1E01 (I.Ed)	17.511	7.677	0.438	15.53	14	29.333
*1F01 (Math)	17.421	8.872	0.509	16.25	18	29.000
*1G01 (Sci)	14.673	8.125	0.553	12.87	15	33.333
*1G02 (Sci)	19.661	8.421	0.428	17.90	18	25.454
*1H01 (S.St)	21.348	9.811	0.459	19.13	17	24.133
*1H02 (S.St)	22.578	10.006	0.443	21.14	23	22.000
*2B01 (Comm)	15.652	8.223	0.525	14.13	17	31.909
*2B02 (Comm)	19.417	10.532	0.542	17.07	14	25.714
*2C01 (Eng)	17.762	12.041	0.678	14.60	8	27.145
*2C02 (Eng)	19.101	14.070	0.736	15.40	12	25.428
*2C03 (Eng)	25.429	18.775	0.738	21.20	15	19.800
*2C04 (Eng)	16.057	14.666	0.913	11.20	4	31.400
*2D01 (H.Ec)	20.114	9.761	0.485	17.90	15	24.142
*2D02 (H.Ec)	19.002	8.828	0.464	17.20	17	25.727
*2D03 (H.Ec)	17.451	8.095	0.463	16.40	17	28.357
*2D04 (H.Ec)	18.071	8.379	0.463	16.80	18	29.500
*2D05 (H.Ec)	14.693	6.565	0.446	13.40	13	34.777
*2E01 (I.Ed)	14.194	6.972	0.491	12.40	12	32.920
*2E02 (I.Ed)	16.402	9.069	0.553	18.50	9	28.562
*2E03 (I.Ed)	18.037	8.744	0.484	16.20	15	28.360
*2F01 (Math)	14.406	6.802	0.472	13.20	9	35.857
*2G01 (Sci)	16.828	8.905	0.529	14.90	15	30.846
*2G02 (Sci)	19.472	9.338	0.479	17.70	11	25.818
*2H01 (S.St)	21.822	12.143	0.556	19.40	21	25.250
*2H02 (S.St)	20.214	9.262	0.458	17.40	16	25.400
*3B01 (Comm)	17.214	7.936	0.461	15.50	10	29.428
*3B02 (Comm)	18.917	8.021	0.424	17.00	18	24.111
*3C01 (Eng.)	13.701	10.242	0.747	10.10	7	40.666
*3C02 (Eng)	22.226	16.081	0.723	17.00	14	21.000
*3F01 (Math)	13.894	7.781	0.560	12.30	10	36.500
*3G01 (Sci.)	18.636	9.441	0.506	17.25	14	27.117
*3G02 (Sci.)	22.054	12.384	0.561	19.85	22	22.642
*3H01 (S.St)	17.522	7.517	0.429	16.20	21	28.033
*3H02 (S.St)	22.952	10.761	0.468	21.00	20	20.666

A Grade 9 English text has the lowest average sentence length and a Grade 10 English text the highest.

One final observation should be made about the sentence length characteristics presented in TABLES XVIII through XXIX. Sentence length and variability are relatively consistent within the three grade level corpora. However, considerable range is evident in average sentence lengths across the samples when organized by grades, subjects across grades, subjects within grades, and by individual textbooks. In addition to this diversity, a striking feature is the considerable variability of the sentence length patterns as indicated by the standard deviations and coefficients of variation reported for the samples when organized by subjects across grades, subjects within grades, and textbooks.

For example, in TABLE XIX for the samples organized by subjects across grades, the standard deviation for the sentence lengths range from 8.150 for Mathematics to 13.685 for English. For the Math samples, approximately 68 percent of the sentences would range from 6.097 to 23.397 words in length with an average of 15.247. For the English samples, 68 percent of the sentences would range from 3.883 to 31.253 words in length with a mean of 17.568.

This variability exists throughout the range of samples, with the exception of grades, and is also evident in the ranges reported for the coefficient of variation and to some degree for the ranges reported for average numbers of sentences per 500 word sample. The coefficient of variation indicates the rate at which items move away from the mean, and the lower the variation the greater the degree of sentence length homogeny in the sample. For example, for the samples organized by subjects across grades (TABLE XIX), the coefficients of variation are quite varied. The coefficient of 0.779 for English indicates that the samples in this subject area are less alike than those in Social Studies which has a coefficient of 0.479. English has, overwhelmingly, the largest coefficient of variation within all the subject and text corpora.

The results of Pearson's skew factor for the sentence length characteristics for the Corpus and various corpora are presented in TABLES XXV through XXVII. A result of zero indicates sentence lengths approximating a normal distribution where the mean and mode coincide. A positively skewed distribution indicates a tailing off to longer sentences while a negatively skewed distribution indicates a tailing off to shorter sentences in relation to the mean. A normal distribution indicates a generally equivalent distribution of long and short sentences about the mean.

The areas most closely approximating a normal distribution of sentence lengths were the Corpus (0.029), Grade 8 (0.060), Grade 8 Mathematics (0.065), Grade 9 Commerce (0.057), three Grade 8 textbook corpora (Mathematics, -0.065; Science, -0.040; Social Studies, 0.042), and a Grade 10 Science textbook with the closest figure of all (-0.004).

The corpora which had the most skewed distributions included Grade 10 (0.770), English (0.772), Grade 9 English (0.737), Grade 9 Mathematics (0.794), six Grade 9 textbooks (English 0.811 and 0.822; Home Economics 0.847; Industrial Education 0.816; Mathematics 0.794; and Science 0.907), and a Grade 10 Commerce textbook, 0.909.

TABLE XXV

PEARSON'S SKEW FACTOR FOR EACH GRADE LEVEL, THE CORPUS AND SUBJECTS ACROSS THE CORPUS

Grade	Skew
8	0.060
9	0.275
10	0.770
Corpus	0.029
Commerce	0.525
English	0.772
Home Economics	0.286
Industrial Educ.	0.672
Mathematics	0.153
Science	0.357
Social Studies	-0.107

TABLE XXVI

PEARSON'S SKEW FACTOR FOR SUBJECTS IN EACH GRADE LEVEL

Subject	Grade 8	Grade 9	Grade 10
Commerce	-	0.057	0.509
English	0.429	0.737	0.681
Home Economics	0.423	0.251	-
Industrial Educ.	0.457	0.179	-
Mathematics	-0.065	0.794	0.500
Science	0.241	0.318	0.186
Social Studies	-0.130	0.557	-0.258
Corpus	0.060	0.275	0.770

TABLE XXVII

PEARSON'S SKEW FACTOR FOR EACH TEXT

Text	Subject	Skew
*1C01	English	0.254
*1C02	English	0.411
*1D01	Home Economics	0.423
*1E01	Industrial Education	0.457
*1F01	Mathematics	-0.065
*1G01	Science	-0.040
*1G02	Science	0.197
*1H01	Social Studies	0.443
*1H02	Social Studies	-0.042
*2B01	Commerce	-0.164
*2B02	Commerce	0.514
*2C01	English	0.811
*2C02	English	0.504
*2C03	English	0.555
*2C04	English	0.822
*2D01	Home Economics	0.523
*2D02	Home Economics	0.226
*2D03	Home Economics	0.557
*2D04	Home Economics	0.847
*2D05	Home Economics	0.257
*2E01	Industrial Education	0.314
*2E02	Industrial Education	0.816
*2E03	Industrial Education	0.347
*2F01	Mathematics	0.794
*2G01	Science	0.205
*2G02	Science	0.907
*2H01	Social Studies	0.677
*2H02	Social Studies	0.455
*3B01	Commerce	0.909
*3B02	Commerce	0.114
*3C01	English	0.654
*3C02	English	0.511
*3F01	Mathematics	0.500
*3G01	Science	0.490
*3G02	Science	0.004
*3H01	Social Studies	-0.462
*3H02	Social Studies	0.274

Task 6.2 To illustrate the data for sentence lengths, sixty-six graphs were produced by the U.B.C. plotting package using a CALCOMP Drum Plotter. These are presented in APPENDIX H. The narrow range of standard deviations for the sentence lengths of the Corpus and most of the corpora is indicated by the leptokurtic nature of these graphs, (sentences tend to cluster around the mean length). The greater degree of variation of sentence lengths in some corpora (English for example), is indicated by the mesokurtic plateau of their graphs.

The graphs provide good visual illustration of the relative distribution of short and long sentences in the sixty-six corpora.

Task 6.3 The repeat-rate frequency tables for the Corpus and corpora are presented in volumes C.V., G.V., S.V., S.G.V., and T.V. (See Task 5.2). The results of Yule's K (for sentences) are listed in TABLES XXVIII through XXX. High K values indicate a greater concentration of commonly occurring sentence lengths while low values indicate a concentration of less frequently occurring sentence lengths.

The K values for each of the three grade levels, the Corpus, and the subject areas across grade levels are presented in TABLE XXVIII. Grade 8 has the smallest value (326.67), indicating a greater variety of sentence lengths used than in Grades 9 and 10.

TABLE XXVIII

K FACTORS (SENTENCES) FOR EACH GRADE LEVEL, THE CORPUS, AND SUBJECTS ACROSS THE CORPUS

Grade	K Factor
8	326.67
9	344.88
10	334.55
Corpus	336.35
Commerce	364.57
English	296.64
Home Ec.	361.32
Ind. Ed.	399.64
Math.	402.07
Science	334.32
Soc. St.	333.49

The great diversity of the K factor in the various subject areas within the grade levels is shown in TABLE XXIX.

TABLE XXIX

K FACTORS (SENTENCES) FOR SUBJECTS WITHIN GRADE LEVELS

Subject	Grade 8	Grade 9	Grade 10
Commerce	-	357.26	397.36
English	279.02	287.23	360.10
Home Ec.	377.56	359.24	-
Ind. Ed.	434.17	398.78	-
Math.	385.35	465.07	427.39
Science	356.30	343.89	319.42
Soc. St.	312.18	291.12	361.56
Corpus	326.67	344.88	334.55

The K factors, range from 279.02 in Grade 8 English to 465.07 in Grade 9 Mathematics.

TABLE XXX presents the K factors for each textbook used in the study. These results range from 204.57 (Grade 9 English) to 484.43 for a Grade 9 Industrial Education textbook.

TABLE XXX

K FACTORS (SENTENCES) FOR EACH TEXT

Text	Subject	K Factor
*1C01	English	244.22
*1C02	English	297.57
*1D01	Home Economics	377.56
*1E01	Industrial Education	434.17
*1F01	Mathematics	385.35
*1G01	Science	403.33
*1G02	Science	348.21
*1H01	Social Studies	311.35
*1H02	Social Studies	308.65
*2B01	Commerce	406.81
*2B02	Commerce	324.85
*2C01	English	310.80
*2C02	English	275.85
*2C03	English	204.57
*2C04	English	337.54
*2D01	Home Economics	335.97
*2D02	Home Economics	344.99
*2C03	Home Economics	368.25
*2D04	Home Economics	351.62
*2D05	Home Economics	449.12
*2E01	Industrial Education	484.43
*2E02	Industrial Education	429.83
*2E03	Industrial Education	365.36
*2F01	Mathematics	465.07
*2G01	Science	362.81
*2G02	Science	342.44
*2H01	Social Studies	266.15
*2H02	Social Studies	311.16

TABLE XXX (CONT.)

K FACTORS (SENTENCES) FOR EACH TEXT

Text	Subject	K Factor
*3B01	Commerce	382.22
*3B02	Commerce	409.44
*3C01	English	400.35
*3C02	English	226.76
*3F01	Mathematics	427.39
*3G01	Science	348.48
*3G02	Science	285.01
*3H01	Social Studies	385.76
*3H02	Social Studies	311.85

English with five of the first ten textbooks ranked, has the greatest number of textbooks with a low value of K. Social Studies has four textbooks out of the first ten and there is one Science textbook ranked number six. Four of the first ten textbooks with low K values are in Grade 8, four are in Grade 9, and two are in Grade 10.

Industrial Education with three of the last six texts and Mathematics with two of the last six texts are the two subjects with the greatest number of high K values. One Home Economics text also had a high K value.

Task 7.

Generate comparative and statistical analyses of the distribution of the 100 most frequently occurring word-types of the Corpus across the three grade levels, the seven subject areas, and the subject areas within the three grade levels.

7.1 Test the following null hypotheses.

Hypothesis 1. There are no significant differences in the actual distribution of the 100 most frequent word-types of the Corpus when compared to the expected distribution of each word-type for:

1.1 the three grade levels of the Corpus,
1.2 the seven subject areas of the Corpus,
1.3 the subject areas within Grade 8,
1.4 the subject areas within Grade 9,
1.5 the subject areas within Grade 10.

7.2 Investigate and describe the number of word-types which differed significantly in their distribution across each of the areas tested in Task 7.1.

Task 7.1 In this analysis, the 100 most frequently occurring words in the total Corpus were used as the basis of comparison. The basic task was to answer the question, "Do the 100 most frequently occurring words derived from the total Corpus have similar frequencies of occurrence when the samples are organized by grade level, subjects across grades, and subjects within grades?" Acceptance of the null hypotheses would indicate that there is substantial similarity (in terms of frequency of occurrence) between the list of the 100 most frequently occurring words in the Corpus and the most frequently occurring words for the various corpora. Chi-square tests were not computed for the thirty-seven textbooks but it would have been possible to do so. A total of 500 chi-squares were computed. Complete data for the chi-square analyses are available in APPENDIX I.

TABLE XXXI provides a summary of the chi-square results. Only two words have similar frequencies of occurrence across all corpora - "as" and "very". The other ninety-eight words exhibit considerable variation in their frequency of occurrence across the various corpora. In all, the null hypothesis was rejected in a total of 372 out of 500 tests, with the level of significance set at .01.

TABLE XXXI

CHI-SQUARE ANALYSIS OF THE 100 MOST FREQUENT WORD-TYPES IN THE CORPUS ACROSS GRADES, SUBJECTS, AND SUBJECTS WITHIN GRADES

Rank	Word	Grades	Subjects (C)	Subjects (8)	Subjects (9)	Subjects (10)
1	THE	-	**	**	**	**
2	OF	**	**	**	**	**
3	AND	**	**	**	**	**
4	A	**	**	**	**	**
5	TO	-	**	**	**	**
6	IN	**	**	-	-	**
7	IS	**	**	**	**	**
8	THAT	**	**	**	**	**
9	IT	-	**	**	-	**
10	ARE	**	**	**	**	**
11	FOR	**	**	**	**	**
12	YOU	**	**	**	**	**
13	BE	**	**	**	**	**
14	AS	-	-	-	-	-
15	OR	**	**	**	**	**
16	WITH	**	**	-	-	**
17	ON	-	**	**	**	-
18	THIS	**	**	**	**	**
19	BY	-	**	**	**	-
20	WAS	**	**	**	**	**
21	HE	**	**	**	**	**
22	FROM	-	**	**	**	**
23	HAVE	**	**	**	**	**
24	AT	-	**	**	**	**
25	WHICH	-	**	**	**	**
26	ONE	-	**	**	-	**
27	NOT	-	**	-	**	**
28	CAN	**	**	**	**	**
29	YOUR	**	**	**	**	**

TABLE XXXI (CONT.)

CHI-SQUARE ANALYSIS OF THE 100 MOST FREQUENT WORD-TYPES IN THE CORPUS ACROSS GRADES, SUBJECTS, AND SUBJECTS WITHIN GRADES

Rank	Word	Grades	Subjects (C)	Subjects (8)	Subjects (9)	Subjects (10)
30	They	**	**	**	**	**
31	We	**	**	**	**	**
32	His	-	**	**	**	**
33	Will	**	**	**	**	**
34	If	**	**	**	**	**
35	An	-	-	**	-	-
36	When	**	**	-	-	**
37	All	-	**	-	**	**
38	But	-	**	-	**	**
39	These	-	**	**	**	**
40	May	**	**	**	**	**
41	There	-	**	**	**	**
42	Has	-	**	**	-	**
43	I	**	**	**	**	**
44	Other	-	**	-	-	-
45	Some	-	**	-	**	-
46	More	-	**	-	**	-
47	Where	**	**	**	-	**
48	Had	**	**	**	**	**
49	Their	**	**	**	**	**
50	Used	**	**	**	**	**
51	Many	-	**	**	**	**
52	So	**	**	**	**	**
53	Each	-	**	**	-	**
54	Two	-	**	**	**	**
55	About	-	**	**	**	**
56	Should	**	**	**	**	**
57	What	-	**	**	**	**
58	Than	-	-	-	**	**
59	Been	**	**	-	-	-
60	Into	-	**	**	-	-
61	Them	-	**	**	**	**
62	Use	**	**	**	**	**
63	Make	**	**	**	**	-
64	Do	-	**	**	**	**
65	Up	-	**	-	**	**
66	Such	-	-	-	**	-
67	Then	**	**	**	-	**
68	Time	-	**	-	-	**
69	Its	**	**	**	**	**
70	Would	-	**	-	**	**
71	How	-	**	**	**	**
72	Number	**	**	**	**	**
73	Made	**	**	**	**	-
74	Out	-	**	-	**	**
75	Most	-	**	-	-	**
76	Only	**	**	**	-	**

TABLE XXXI (CONT.)

CHI-SQUARE ANALYSIS OF THE 100 MOST FREQUENT WORD-TYPES IN THE CORPUS ACROSS GRADES, SUBJECTS, AND SUBJECTS WITHIN GRADES

Rank	Word	Grades	Subjects (C)	Subjects (8)	Subjects (9)	Subjects (10)
77	No	-	**	-	-	-
78	Must	**	**	-	**	**
79	Water	**	**	**	**	**
80	Also	-	**	**	**	-
81	First	-	-	-	**	-
82	Very	-	-	-	-	-
83	Good	**	**	**	-	**
84	Him	-	**	**	**	**
85	Same	-	**	**	**	**
86	Could	**	**	**	**	**
87	Who	-	**	**	**	**
88	Any	-	**	**	**	**
89	Because	-	**	**	-	**
90	See	-	**	**	-	**
91	Like	-	**	-	**	**
92	Much	-	**	-	**	**
93	People	-	**	**	-	**
94	Called	-	**	**	**	**
95	Place	**	**	**	**	-
96	Through	-	**	-	**	**
97	Work	**	**	**	**	-
98	New	**	**	-	**	**
99	Small	-	**	-	-	**
100	Over	-	**	-	**	-

** SIGNIFICANT AT THE .01 LEVEL.
- NOT SIGNIFICANT.

<u>Task 7.2</u> A breakdown of the chi-square results for grades, subjects across grades, and subjects within grades is presented in TABLE XXXII. The greatest similarity in the distribution of commonly occurring words appears when the samples are organized by grades. Only 46 out of 100 chi-square tests were rejected. In the case of subjects across grades, 94 chi-square tests were rejected. Similar results are evident for subjects within Grades 8, 9, and 10 with 72, 76, and 81 chi-square tests rejected.

These results are also shown in the pattern of rejection in TABLE XXXI.

TABLE XXXII

SUMMARY OF CHI-SQUARE ANALYSIS OF 100 MOST FREQUENT WORD-TYPES IN THE CORPUS ACROSS GRADES, SUBJECTS, AND SUBJECTS WITHIN GRADES

	Gdes in Corpus	Subj. in Corpus	Subj. in Gde 8	Subj. in Gde 9	Subj. in Gde 10
Sig.	46	94	72	76	81
Non Sig.	54	6	28	24	19
Total	100	100	100	100	100
d.f	2	6	5	6	4

Task 8.

Do the sentence length distributions of the three grade level corpora , the seven subject area corpora, and the eighteen subject within grade level corpora differ from the sentence length distribution of the Corpus? This task involved testing the following null hypotheses.

Hypothesis 2. There are no significant differences in the actual distribution of short, average, and long sentences when compared to the expected distribution of each of the sentence lengths for:

2.1 the three grade levels of the Corpus,
2.2 the seven subject areas of the Corpus,
2.3 the subject area corpora within Grade 8,
2.4 the subject area corpora within Grade 9,
2.5 the subject area corpora within Grade 10.

Task 8. The purpose of this analysis was to determine if the sentence length distributions were similar across the various corpora. It would have been unwieldly to determine the similarity in the distributions of all sentence length types across all the corpora involved. For example, 94 different sentence lengths were identified for the Corpus alone, ranging in length from one word to 117 words. The decision was made to select a number of sentence lengths representative of the Corpus as a whole and determine if these exhibited similar distributions across the various corpora. The sentence length distribution for the total Corpus (See FIGURE 7.1, APPENDIX H) was carefully scrutinized and five sentence lengths selected as being representative of the Corpus on the basis of their relative frequency of occurrence were chosen. The five sentence lengths represent a group on either side of the Corpus mean (10 and 20 words in length respectively), plus three groups of larger sentences (30, 40 and 50+ words in length respectively). The sentences of 50+ words represent all the larger sentences in the positively skewed distribution for sentence lengths. This end of the curve represents the small quantities and great varieties of sentences 50 words and over in length.

The basic task was to answer the question, "Do the five sentence lengths derived from the Corpus have similar distributions across the corpora when the samples are organized by grade levels, subjects across grades, and subjects within grades?" Acceptance of the null hypothesis would indicate that there is substantial similarity between the distribution of the representative sentence lengths of the Corpus and the distribution of sentences across the various corpora. Chi-square tests were not computed for the thirty-seven textbooks but it would have been possible to do so. Complete data for the chi-square analyses for the five sentence lengths across the various corpora are available in APPENDIX J. A total of twenty-five chi-square tests were computed.

TABLE XXXIII provides a summary of the chi-square results. In all tests the null hypothesis was rejected at the .01 level of significance illustrating the diversity in the sentence length distributions for the representative sentences by grades, subjects across grades, and subjects within grades. It should be pointed out that there is greater apparent similarity in the sentence length distributions for the sample organized by grades than when they are organized by subjects across grades or subjects within grades.

TABLE XXXIII

CHI-SQUARE ANALYSIS OF SELECTED SENTENCE LENGTHS FOR THE GRADES, SUBJECTS ACROSS GRADES, AND SUBJECTS WITHIN GRADES

	Grades	Subjects Corpus	Subj. in Gde 8	Subj. in Gde 9	Subj. in Gde 10
X^2 value	21.98	152.23	53.33	109.63	41.68
.01 level	18.48	42.98	37.57	42.98	32.00
d.f	8	24	20	24	16

<u>Task 9.</u>
Develop an "elimination technique" for selecting the most significant content words in a word list using the ranked frequency lists developed for the Corpus, the three grade level corpora, and the subject area corpora.

9.1 Produce a set of graphs to illustrate the word frequency by rank of the Corpus, the three grade level corpora, and the seven subject area corpora.

9.2 What is the effect of eliminating the highest frequency words and the lowest frequency words from the total spectrum of words for each of the areas stated in 9.1?

9.3 Can the residual of words remaining after eliminating the high and low frequency words described in 9.2 serve as a pool for selecting the most useful content words for the Corpus, the three grade level corpora, and the seven subject area corpora, through analyses based on relative frequency of occurrence and subjective criteria?

Task 9.1 This task was designed to determine the feasibility of developing an "elimination technique" for selecting the most significant vocabulary from a list of words derived from samples of natural language text representative of prescribed subject materials. The graphs illustrating the word frequency distributions for the Corpus, grade level, and subjects across grades corpora used in this task are presented in APPENDIX K. The graphs for subjects within grades and the thirty-seven textbooks were not plotted although it would be possible to do so.

The graphs approximate the shape usually found in the analysis of word frequency data. The graphs for the grade level and subject corpora have the same general shape as the word frequency distribution for the Corpus. The Corpus graph (see FIGURE 5) illustrates the high frequency for the first 100 most common words, the clustering of word frequencies about the mid point of the graph, the gradual tailing off to words occurring three times or less, and the final tabulation of the *hapax legomena* .

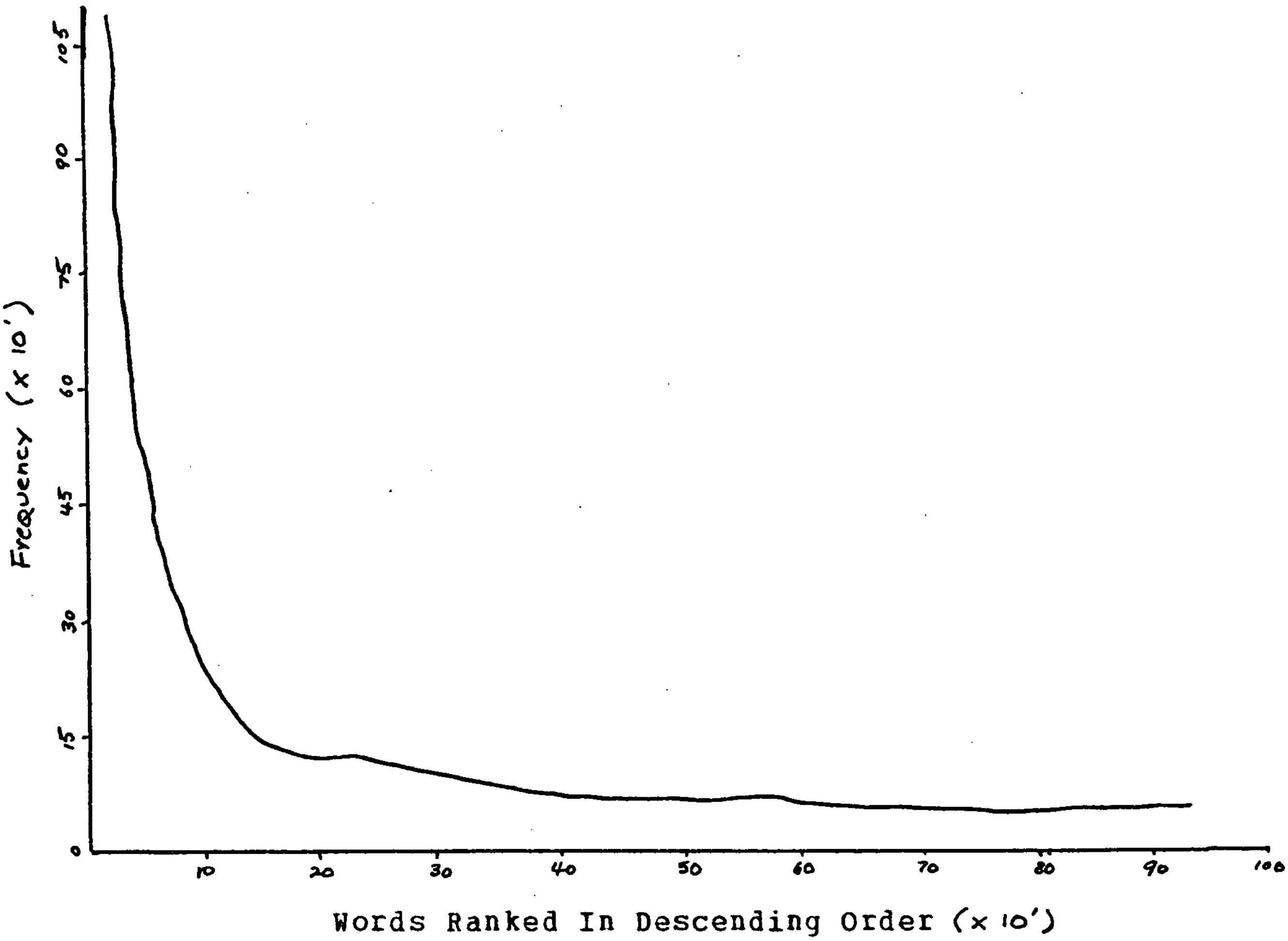


FIGURE 5

WORD FREQUENCY DIAGRAM OF THE CORPUS

Task 9.2 The word frequency graph of the Corpus (FIGURE 5) is used to illustrate the effect of applying the "elimination technique" to a word list. Point A represents the cutoff point for 50 percent of the high frequency tokens in the Corpus and Point B represents the cutoff point for 10 percent of the low frequency tokens in the Corpus. The remaining 40 percent of tokens between points A and B are considered to represent the "significant" body of content for the Corpus. The distribution for these words would most likely approximate a normal curve with a mean frequency of occurrence and proportionate tailing off from both sides of the mean.

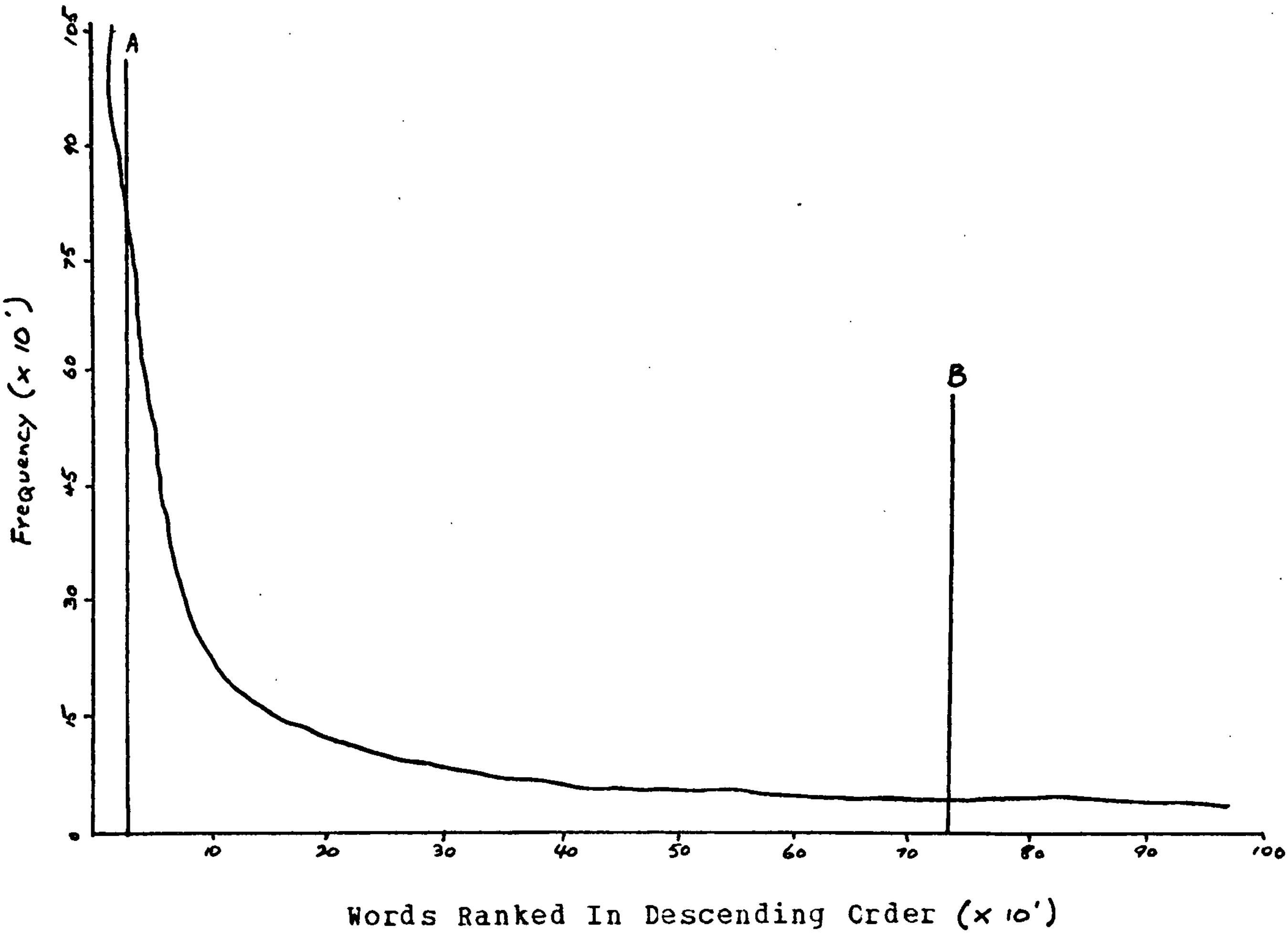


FIGURE 6

APPLICATION OF THE "ELIMINATION TECHNIQUE" TO THE WORD FREQUENCY DIAGRAM OF THE CORPUS

TABLE XXXIV presents the data for determining the number of word-types and the percentage of the total word-types accounted for by both the cutoff points A and B in the Corpus, the three grade level corpora, and the seven subject area corpora.

TABLE XXXIV

NUMBER AND PERCENTAGE OF WORD-TYPES ELIMINATED BY POINT A (50% CUTOFF OF TOKENS) AND POINT B (10% CUTOFF OF TOKENS)

	Point A		Point B		
	No.of Word Types	% of Word Types	No.of Word Types	% of Word Types	Total No.of Word Types
Corpus	111	0.68	12,695	77.40	16,405
Grade 8	94	1.33	4,593	65.36	7,027
Grade 9	109	0.96	7,730	67.80	11,401
Grade 10	118	1.52	4,906	63.40	7,736
Commerce	82	2.72	1,904	63.00	3,020
English	92	1.30	4,010	72.60	7,079
Home Economics	81	1.47	3,832	69.30	5,529
Industrial Education	90	2.22	2,433	60.00	4,060
Mathematics	53	2.72	1,298	66.50	1.952
Science	96	1.99	2,968	61.41	4,833
Social Studies	120	1.93	3,144	50.62	6,211

The words up to Point A account for a very small number of word-types in each of the eleven distributions. The Corpus, which had 111 word-types 'eliminated', represented the smallest percentage (0.68), while Commerce with 82 word-types and Mathematics with 53 word-types 'eliminated' respectively, both had 2.72 percent of their total word-types deleted.

The great majority of the word-types which would be deleted from the distribution using this technique would be high frequency structure words which are similarly common to the

Corpus and the ten corpora being investigated. These words, which constitute 'noise in the system', are not considered distinct enough to have special significance for the content material they represent.

The words up to Point B account for the majority of word-types in each of the eleven distributions. The numbers of word-types 'eliminated' ranged from a low of 3,144 (50.62 percent) in Social Studies, to a high of 12,695 (77.4 percent) in the Corpus.

Most of the low frequency words deleted would be items which occur only several times in a distribution. These words are considered to be too rare to have special significance for their respective content materials.

The complete listing of the word-types and tokens for each of the eleven areas is provided in the following volumes available from the Computing Centre, at the University of British Columbia. The organization of these volumes was described previously under Task 4.

1) Corpus (Volume C.V.)

2) Grade Levels (Volume G.V.)

3) Subject Areas (Volume S.V.)

<u>Task 9.3</u> The balance of the words remaining between points A and B (approximately 40 percent of tokens) in each of the corpora are considered to be neither too common nor too rare and

have the greatest significance for the content material they represent. TABLE XXXV presents the number and percentage of word-types between the A and B cutoff points for the Corpus, grades, and subjects across grades corpora. The vast majority of word-types in this section of the distributions are lexical items which occur seven times or more in the Corpus, and three times or more in most of the corpora.

TABLE XXXV

NUMBER AND PERCENTAGE OF WORD-TYPES
BETWEEN POINT A AND POINT B (40% OF TOKENS) FOR THE CORPUS, GRADES, AND SUBJECTS ACROSS GRADES

	No. of Word-Types	% of Word-Types	No. of Word-Types
Corpus	3,599	21.92	16,405
Grade 8	2,340	33.31	7,027
Grade 9	3,562	31.24	11,401
Grade 10	2,712	35.08	7,736
Commerce	1,034	34,28	3,020
English	2,977	26.10	7,079
Home Economics	1,616	29.23	5,529
Industrial Education	1,537	37.78	4,060
Mathematics	601	30.78	1,952
Science	1,769	36.60	4,833
Social Studies	2,947	47.45	6,211

It is interesting to note that the really significant lexical content, once the common words and the rarely occurring words are 'eliminated', consists of a relatively small number of word-types when compared to the total for the complete sample organized by Corpus, grades, and subjects across grades. Further subjective evaluation of the words remaining between points A and B by subject specialists would no doubt reduce the total even further.

SUMMARY

This chapter has presented the analysis of the data and the findings for the study. Nine tasks were involved and the completion of the tasks resulted in the production of some 5,500 pages of printed material which included facsimiles of the instructional materials sampled and the sixty-six word lists plus accompanying tables, graphs, and statistical summaries.

Task 1 A representative sample of instructional materials was selected and organized. The Corpus consisted of 469, five hundred word samples of natural language taken from thirty-seven prescribed English language textbooks representing seven subject areas.

Task 2 The Corpus was keypunched onto IBM computer cards using the FMT computer program and stored on disk to await processing.

Task 3 Two editions of the Corpus were produced. One edition was organized by grade levels and one organized by subject areas. This enabled the production of an additional sixty-five corpora with the samples organized by the three grade levels, seven subject areas, eighteen subjects within grade levels, and thirty-seven textbooks.

Task 4 Two word frequency lists, one organized alphabetically and one organized by descending rank order, were produced for the Corpus and the sixty-five corpora. Tables presenting the rank of word frequency figures in descending and ascending order were developed for each of the sixty-six corpora.

Task 5. Comparative and statistical analyses were generated based on the lexical characteristics of the Corpus and the sixty-five corpora. Yule's characteristic K was computed to illustrate the concentration of commonly occurring vocabulary in each of the sixty-six corpora.

The heaviest load of new reading material as measured by total tokens was introduced in Grade 9 which also has the heaviest loading of specific word-types in the three grades. Home Economics and English were the two largest subject corpora when the samples were organized across the three grade levels by subjects, with English having a much higher proportion of word-types suggesting a greater vocabulary load. Social Studies also had a high proportion of word-types.

When the samples were organized by subjects within grades, Home Economics and Social Studies have the largest subject corpora and Social Studies and English have the largest vocabulary load for Grade 8; Home Economics and Industrial Education have the largest subject corpora and Home Economics and English have the largest vocabulary load for Grade 9; and, Science and Social Studies have the largest subject corpora and also the largest vocabulary load for Grade 10.

With the samples organized by textbooks, a Grade 10 Social

Studies text and a Grade 9 Industrial Education text have the largest subject corpora while the same Grade 10 Social Studies text and a Grade 9 English text have the largest vocabulary load.

Thus, it is evident that there is considerable diversity in word-type and token distribution when the samples are organized by the various grade, subject and text corpora.

The application of the K characteristic to measure density of commonly occurring vocabulary again indicated that Grade 9 had the lowest K value and that Home Economics and English had the greatest variety of vocabulary used across the various corpora except for subjects within Grade 10 where Commerce and English have the lowest density of commonly occurring words. With the samples organized by textbooks, English texts consistently have the lowest K values.

Task 6. Comparative and statistical analyses were generated based on the sentence length distributions of the Corpus and the sixty-five corpora. Graphs for each of the sixty-six corpora were developed for this task. Yule's K characteristic was used to describe the concentration of commonly occurring sentence length types in the Corpus and sixty-five corpora.

Relatively uniform average sentence lengths were found when the samples were organized by grades. This pattern was also repeated when the samples were organized by the subjects across the three grades although the range in averages increased.

When the samples were organized by subjects within Grade 9, fairly uniform average sentence lengths are evident with the exception of Social Studies. However, considerable variability in the sentence length distributions are evident as indicated by the range in standard deviation, coefficient of variation and to some extent by the average sentence lengths reported per 500 word samples. With the samples organized by subjects within Grade 10, the same pattern is exhibited with the exception that Science has the largest average sentence length. With the samples organized by textbooks, a Grade 10 English text has the largest average sentence length.

It should be pointed out that with the samples organized by subjects across grades, subjects within Grades 8, 9, and 10, and by textbooks, English exhibits the greatest variation in sentence length distribution. No other subject area approaches the magnitude of the standard deviation and coefficient of variation reported for English samples.

The application of Yule's K characteristic to measure the density of commonly occurring sentence lengths indicated that Grade 8 had the lowest K value; English had the lowest K value for samples organized by subjects across grades; English had the lowest K value for samples 8 and 9; Science had the lowest K value for Grade 10; and two English texts had the lowest K values for the samples organized by textbooks.

Task 7. Chi-square tests were computed to illustrate the distribution of the 100 most commonly occurring word-types of the Corpus across the three grade level corpora, the seven subject area corpora, and the eighteen subjects within grade level corpora.

A total of 500 chi-square tests were computed for the samples organized by grade level, subjects across grades, and subjects within grades corpora to determine the nature of the distribution of the 100 most common words. The results indicated that there was significantly more variability in the use of the most commonly occurring vocabulary when the samples were organized by subjects across grades and subjects within grades than when they were organized by Grades 8, 9, and 10.

Task 8. Chi-square tests were computed to illustrate the distribution of five selected sentence lengths of the Corpus across the three grade level corpora, the seven subject area corpora, and the eighteen subjects within grade level corpora.

The results of the chi-square tests on the selected sentence lengths for the various corpora indicated that significant diversity exists in sentence length distribution for even the most common sentence lengths across the corpora when the samples are organized by grades, subjects across grades, and subjects within grades. An analysis of the chi-square results within a particular test suggested that there was greater apparent similarity in the sentence length distributions organized by grade levels than by subject areas or subjects within grades.

Task 9. An "elimination technique" was described for use in selecting the most "significant" content words in the word lists of the Corpus, the three grade level corpora, and the seven subject area corpora. A set of graphs was constructed to represent the word frequency by rank of each of the eleven areas investigated.

The use of the "elimination technique" to determine the "significant" words in a body of print material illustrated the great influence a relatively small number of highly frequent structure words have on the word frequency distibution. Once these common words had been 'eliminated', along with a number of very rare words, a core of 'significant' vocabulary is available for examination and analysis.

Computers and Language Analysis

Computer techniques were used extensively in most aspects of the study. Over 200 specially prepared computer files and twenty computer programs were developed to generate a Corpus of 235,107 words; sixty-five smaller corpora representing grade levels, subject areas, subject areas within within grades, and textbooks; tables, figures, and graphs; numerous statistical procedures used to analyze the material; and, print the final copy of the dissertation itself.

The data generated by the study were formatted using the FMT program and occupied over 5,500 pages which were organized into eight volumes. All information pertaining to the study was placed on magnetic tapes and stored in the Computing Centre and the Special Collections Division of the Library at the

University of British Columbia.

The 3270 Conversational Terminal (CRT) was used throughout the study to monitor the input of data, edit the material and organize the production of the Corpus, corpora, word lists and accompanying statistics.

Apart from the Corpus, which contained nearly a quarter of a million words, considerable reorganization was required to develop the other sixty-five corpora used in the study. The magnitude of the file management task was further complicated by the need to compile two word lists (one in alphabetical order and one in descending rank order) for the Corpus and for each of the sixty-five corpora, as well as to develop two tables illustrating descending and ascending order for each of the sixty-six corpora used in the study. In addition, it was necessary to program a thorough examination of the word and sentence length characteristics of the Corpus and the corpora, provide relevant graphs and tables, and test the statistical hypotheses.

The use of existing computer techniques, plus the development of new computer programs, enabled the tasks described above to be rapidly facilitated and also provided the necessary statistical information required for the study.

CHAPTER V

DISCUSSION, CONCLUSIONS, AND RECOMMENDATIONS

This chapter presents a discussion of the major findings of the study. A number of conclusions drawn from the findings are given and the relationship of these conclusions to the role of reading in the secondary school discussed. Finally, recommendations for future research resulting from the study are presented.

The central focus of the study involved the use of computer technology to 1) develop a representative Corpus and a series of related corpora of samples of natural language text selected from English language instructional materials prescribed for use in British Columbia junior secondary grades, and 2) make a number of descriptive and comparative analyses of selected word and sentence features of the Corpus, and the grade level, subject area, and textbook corpora.

The study was organized into nine tasks which involved selecting and sampling procedures, methods of data collecting and recording, data processing and analysis, and the posing of relevant research questions to be answered and null hypotheses to be tested. A Pilot Study was conducted prior to the commencement of the investigation to validate sampling techniques and methodological procedures, and generate needed computer programs.

DISCUSSION OF MAIN FINDINGS

The detailed findings of the study are presented in Chapter IV. They are discussed here under the headings: Sampling and Processing Procedures, Production of the Corpus, Production of Word Lists, Lexical Characteristics, Sentence Characteristics, Common Words, Selected Sentence Lengths, and Elimination Technique.

Tasks 1 and 2: Sampling and Processing Procedures

The 235,107 word Corpus derived from the 469, five hundred word samples taken from the thirty-seven textbooks was developed and prepared for computer processing.

The total sample used in the study provided an adequate data base for the various descriptive, comparative, and statistical analyses performed on the Corpus and sixty-five corpora. The FMT computer program was an ideal instrument for the processing of the natural language samples used in the investigation.

Task 3: Production of the Corpus

Two copies of the Corpus, one organized by grade levels

(C.G.) and the other organized by subject areas (C.S.), were produced.

The organization of the two editions of the Corpus (one arranged by grade levels and the other arranged by subject areas) provided useful access to the 469 samples used in the study.

Task 4: Production of Word Lists

The various samples were organized into alphabetical and descending rank order word lists for the Corpus (C.V.), the three grade levels (G.V.), the seven subject areas (S.V.), the eighteen subjects within grades (S.G.V.), and the thirty- seven textbooks (T.V.). Statistical tables for each of the sixty-six corpora described above were also developed.

The development of 132 word frequency lists in both alphabetical and descending rank order along with statistical summary tables facilitated the rapid location of specific word-types and tokens throughout the Corpus and the sixty-five corpora.

Task 5: Lexical Characteristics

The processing of the total Corpus of 235,107 tokens resulted in the identification of 16,405 specific word-types. These results are proportionately similar to the type and token distributions found in other recent research based on computer generated corpora of various sizes including: Kucera and Francis (1967) 50,406 types, 1,014,232 tokens; Carroll, Davis, and Richman (1971) 86,741 types, 5,088,721 tokens; and, Harris and Jacobson (1972) 80,000 types, 4,500,000 tokens.

A pattern was evident in type and token characteristics with the samples organized by grade levels. The corpus of material for Grade 9 was twice the size of that for both Grade 8 and Grade 10 in terms of tokens and 50 percent greater in size in terms of word-types. Nineteen textbooks were used in Grade 9 and nine in Grade 10. This suggests that the middle year in the junior secondary grades exhibits potentially heavier reading demands than either Grade 8 or Grade 10. However, it should be noted that textbooks used in Grade 9 English, Home Economics and Industrial Education can be repeated in Grade 10 and the reading load for this grade depends on the specific use made of these textbooks. With this in mind, one might assume that a marked increase in the quantity of reading content and sheer vocabulary exposure occurs in Grade 9 and most likely continues into Grade 10. Further research would have to be conducted to determine the features of the reading demands in Grade 11 and 12.

With the samples organized by subjects across grades, subjects within grades, and textbooks, no apparent pattern existed in the data except for considerable diversity in word-type and token distribution. However, application of Yule's K characteristic, which provides a statistical indices of the concentration of vocabulary within print materials, resulted in clear trends based on repeat rate frequency for the various grade, subject and text corpora. Grade 9 had the least redundancy in vocabulary of the three grades; Home Economics and English had the least redundancy, to a large degree, in vocabulary in comparison to other subjects within each grade (with Commerce in Grade 9 also exhibiting a low K value); and Home Economics and English had the least redundancy in vocabulary in comparison to other textbooks (with the exception of the low K value for Commerce texts). Considering that K is a measure of the degree to which the token distributions tend to have different words, English and Home Economics clearly make proportionately greater vocabulary demands. The token distributions for all subject word lists also display considerable variation.

These results provide striking evidence for the value of using measures such as the K characteristic to supplement the usual type and token statistics computed in word frequency studies. Determining the specific number of word-types and tokens can provide useful data, but a statistical measure based on the relative redundancy in occurrence of those words allows for sharper differentiation of the real vocabulary demands of various subject areas.

Task 6: Sentence Characteristics

No apparent pattern in sentence length distribution was evident for the samples organized by grades. With the samples organized by subjects across grades, subjects within grades, and textbooks, considerable range of variation in sentence length was apparent. In addition, English overwhelmingly exhibited the largest standard deviations and coefficients of variation in sentence length statistics.

Application of the K characteristic indicated that Grade 8 had the least redundancy in repeat rate of sentence lengths; English had the least redundancy with samples organized by subjects across grades and with the samples organized by subjects within Grades 8 and 9; Science had the least redundancy for Grade 10; and, English had the least redundancy for samples organized by textbooks. English again, as in the case of vocabulary, makes exceptional demands in terms of sentence length variety.

Although English is focused on here because of its rather significant demands in terms of lack of vocabulary redundancy and minimal sentence length repetition, it should be pointed out that with the data available from this study, it is possible to easily develop useful descriptive statements on vocabulary redundancy and sentence length repeat rate for a great variety of configurations for the samples organized by grades, subjects and textbooks.

Task 7: Common Words

The type and percentage of "common words" found to be most frequently represented in the samples of this study are relatively similar to the results obtained in other word count studies. (See for example, the three corpora referred to previously).

The results for the chi-square analyses for samples organized by grades, subjects across grades, and subjects within grades, provide statistical evidence for the assumption that little uniformity exists in the distribution of even the most commonly occurring word-types used in writing. There tended to be a greater uniformity with the samples organized by gross grade groupings than when they were organized by subjects. The style and content characteristics of the separate subject areas are thus significantly instrumental in affecting the frequency of occurrence of even the most common words found in English.

Task 8: Selected Sentence Lengths

The results of the chi-square analysis for samples organized by grades, subjects across grades, and subjects within grades provide statistical evidence for the assumption that little uniformity exists in the distribution of representative sentence lengths. In no subject or grade groupings did the samples follow a homogeneous pattern. There tends to be more uniformity with the samples organized by gross grade groupings than by subjects across and within grades. These results parallel those found for the common word analysis and again suggest that the style and content characteristics of the separate subject areas are also significantly instrumental in affecting the frequency of occurrence of representative sentence lengths.

Task 9: Elimination Technique

An elimination technique was developed, with cutoff points suggested at the 50 percent of the high frequency tokens and 10 percent of the low frequency tokens, using the Corpus word list as a model. This analysis also revealed that the total Corpus and the separate grade and subject corpora contain a larger number of rare word-types even though the large majority of running words are fairly common words. Full comprehension of print sources would involve knowledge of all word-types. However, this is seldom possible and the elimination technique is useful in ascertaining the most significant vocabulary for instructional purposes. The elimination technique (based on the elimination of highly frequent and relatively rare words) could be useful in determining the most significant content in a word list when coupled with the application of judgment by subject specialists.

Word frequency can be justified as a measure of word significance. Certain words are normally repeated as an author develops a topic. When the most significant of the words are separated from words that serve to tie writing together, vocabulary lists with high content significance result. This is particularly true in expository writing where there is little probability that a given word is used to reflect more than one idea.

CONCLUSIONS

In Chapter I the question was asked, "What are the linguistic characteristics of the print sources prescribed for use in Canadian secondary schools?" This study provides partial answers to that question for materials prescribed for the junior secondary grades. The major answer to the question can be stated, "Print sources exhibit extremely diverse characteristics when examined in relation to; quantity of material prescribed, vocabulary redundancy, sentence characteristics, distribution of common words, and the distribution of representative sentence types." In fact, little congruity of pattern exists across the samples of the study when the results are organized to reflect the print sources prescribed for grades, subjects across grades, subjects within grades, and samples by textbooks within the subjects themselves. The variability is marked even in looking at data based on straightforward lexical variables such as word frequency and sentence length.

In all cases, organization of the samples into gross grade patterns masked the subject differences so obvious when the print sources were organized into various combinations representing across and within subject groupings. It would thus be more precise to speak of reading demands in the junior secondary years in terms of subjects across the three grades, subjects within the three grades, or by separate text, rather than by gross grade level alone.

The separate materials in each subject area make unique reading demands as print sources when compared within subjects or to other subject areas within or across grades. Uniformity is lacking in the distribution of even the most common words comprising 50 percent of running prose. The same holds true for the distribution of a representative set of sentence lengths. While there is considerable variation in the vocabulary and sentence style demands in all subjects, the very unique demands of the English genre (and to some extent Home Economics and Commerce) must be pointed out. No other subject area consistently exhibits such variability in vocabulary redundancy, sentence length characteristics, and sentence length homogeny. English materials tend to have a greater concentration of relatively uncommon words over a great variety of sentence lengths. It is assumed here that variability is related to reading difficulty and that widely fluctuating patterns of repeat rate frequency of words and diverse sentence length characteristics are more difficult for the reader to cope with than materials exhibiting a more even distribution of these characteristics.

The results of this study are based on samplings from one print source, prescribed "A" issue texts, and the characteristics of only two relatively straightforward lexical features, word frequency and sentence length, are examined. The variability in the results would possibly be even more pronounced if total samples had been analyzed and if samples from all types of print sources (including supplementary, reference and recreational reading had been included. In addition, if probes were made and statistics developed on a broader array of syntactic and semantic variables related to grammatical functioning, syntax, and logical relationships, a greater variability would be expected.

In conclusion, in describing the reading demands of print materials prescribed for use in junior secondary grades, the variability of the word and sentence characteristics within each subject area are the most obvious factors to be considered. This suggests that realistic reading instruction for secondary schools must focus on the subject areas and the specific print materials used as tools in presenting the ideas and concepts in those subject areas. Such instruction may best be viewed as a shared responsibility between the subject teacher and the reading specialist rather than the sole province of the reading specialist. The subject specialist brings unique knowledge and insight of the discipline and its print sources to the team, while the reading specialist contributes knowledge of the underlying processes and skills of the reading act and familiarity with the characteristics of print in general which contribute to problems in comprehending instructional materials.

RECOMMENDATIONS

This study suggests a number of practical recommendations for the immediate implementation of the main findings and also avenues for future research.

1. The word frequency lists produced for the Corpus and the sixty-five corpora provide subject teachers and coordinators, reading specialists, and school administrators at the junior secondary level with a valuable source of language data

representing each grade level, subject area, subject area within a grade, and individual textbook. The word lists should be examined and their relevance to instruction in regular classroom settings, adult basic education, and classes for New Canadians determined.

2. A number of correlational analyses could be made with the word lists from the present study and word lists previously developed by Lorge-Thorndike, Kucera-Francis, and Carroll et al. This comparison could identify basic differences between data bases compiled from print sources in two different countries and aid in determining the basic differences between Canadian and American English.

3. The Corpus of representative samples generated in the study could provide a useful data base for research in a number of areas. The samples could be used in readability research. For example, it would be relatively easy to generate mutilated samples for Cloze research by developing computer programs to modify the samples and delete every "nth" word. Research could be undertaken to determine the effect of differing sample length and number of samples in the application of existing readability measures. A useful project would be the development of a computer program for syllable counts for application in readability research. The samples themselves could also be further analyzed using techniques and measures from studies in transformational grammar and other linguistic algorithms. Such studies could provide further insight into the role the structure of print materials plays in the processing of written

language.

4. A thorough analysis of the readability of the various textbooks used in the study could be readily undertaken. The 469 samples of approximately 500 words each in length have been carefully selected and described. The data could be added to and updated as new adoptions are made or as the curriculum is revised in subsequent years.

5. An area of research requiring continued attention concerns the different patterns of language in the subject areas. There is a need to further identify what Bormuth (1969) referred to as 'the manipulable linguistic variables which bear a causal relationship to the difficulty of the instructional materials being used'. With this information it would be possible to develop teaching strategies to help students cope with the reading demands presented in their instructional materials.

6. Further analysis should be made into the linguistic characteristics of textbooks within a subject area to determine the specific reading difficulties inherent in certain types of written expression. For example, a textbook dealing with instruction in English expression may offer suggestions on improving sentence construction in one part of the book and a few pages later present a literary excerpt as an example of good writing style.

7. The use of the "elimination technique" could be refined and developed to produce a core vocabulary for each of the

subject areas. These vocabulary lists would provide valuable information in the development of summative, formative, and placement evaluation in reading.

8. Computer techniques should be further developed and modified to allow for further analysis of natural language samples. In addition, a vital need exists for researchers in education to become aware of the advantages of using the computer in their work, to gain an understanding of basic computer procedures, and to communicate their needs and objectives to the computer programmers and other technicians who are available for consultation and advice.

9. The model developed in this study could be modified in a number of ways. Initially, the model could be enlarged to deal with a sample of other textbooks and instructional materials used in the junior secondary grades. This would provide for a wider representation of printed samples and may supply further insights into linguistic variables encountered by students in their reading. Secondly, the model could be extended to encompass Grade 4 to Grade 12, thus enabling a thorough description and analysis to be made of the subject areas within and across the elementary, junior secondary, and senior secondary grade levels. The model could also be applied in other provinces or in studies based on samples across provinces. Finally, the model could be adapted to allow a more detailed analysis of selected linguistic features within a language sample which would provide important information for researchers, subject area teachers, and reading specialists.

BIBLIOGRAPHY

A. BOOKS AND MONOGRAPHS

Alford, M.H.T. "Computer Assistance in Language Learning and in Authorship Indentification". The Computer in Literary and Linguistic Research, ed R.A. Wisbey, London: Cambridge University Press, 1971.

Artley, A Sterl. "The Development of Reading Maturity in High School - Implications of the Gray-Rogers Study". Improving Reading in Secondary Schools: Selected Readings, ed. Lawrence E. Hafner, New York: Macmillan Co., 1967.

______ Trends and Practices in Secondary School Reading. Newark, Delaware: International Reading Association, 1968.

______ "Implementing a Developmental Reading Program on the Secondary Level, Teaching Reading in High School: Selected Articles, ed. Robert Karlin, Indianapolis, New York: Bobbs-Merrill Co. Inc., 1969.

Aukerman, C.R. Reading in the Secondary School Classroom. New York: McGraw-Hill Book Company, 1972.

Ballou, Stephen V. A Model for Theses and Research Papers. Boston: Houghton Mifflin Company, 1970.

Bond, Guy L., and Eva Bond. Developing Reading in High School. New York: MacMillan Co., 1941.

Bond, Guy L., and Miles A. Tinker. Reading Difficulties: Their Diagnosis and Correction. New York: Appleton-Century-Crofts, 1967.

Bormuth, John R. Readability in 1968. A Research Bulletin, National Council of Teachers of English, 1968.

Botel, Morton. Botel Predicting Readability Levels. Chicago: Follett Publishing Co., 1962.

Buckingham, B.R., and E.W.Dolch. A Combined Word List. Boston: Ginn and Co., 1936.

Burton, Dwight L. "Heads Out of the Sand: Secondary Schcols Face the Challenge of Reading". Teaching Reading in High School: Selected Articles, ed. Robert Karlin, Indianapolis, New York: Bobbs-Merrill Co. Inc., 1969.

Campbell, William R. Form and Style in Thesis Writing. Third edition. Boston: Houghton Mifflin Company, 1969.

Carroll, John B. "The Nature of the Reading Process". Theoretical Models and Processes of Reading, ed. Harry Singer, Newark, Delaware: International Reading Association, 1970.

Carroll, John B. "Development of Native Language Skills Beyond the Early Years". The Learning of Language, ed. Carroll E. Reed, New York: Appleton-Century Crofts, 1971.

Carroll, John B., Peter Davies, and Barry Richman. The American Heritage Word Frequency Book. New York: Houghton Mifflin Company, 1971.

Chall, Jean. Readability: An Appraisal of Research and Application. Bureau of Educational Research Monographs, No. 6, Columbus, Ohio: Ohio State University, Bureau Of Educational Research, 1958.

Chomsky, N. Syntactic Structures. The Hague: Mouton, 1957.

______ Aspects of the Theory of Syntax. Cambridge, Mass.: M.I.T. Press, 1965.

Cole, Luella. The Teacher's Handbook of Technical Vocabulary. Bloomington, Illinois: Public School Publishing Company, 1940.

Coleman, E.B. "Experimental Studies of Readability", Readability in 1968, ed. John R. Bormuth, National Council of Teachers of English, 1968.

Coombs, Clyde H. A Theory of Data. New York: John Wiley & Sons, Inc., 1964.

Davis, Frederick B. "Research in Reading in High School and College", Review of Educational Research, 22 (April, 1952), 65-75.

Dechant, E.V. Improving the Teaching of Reading. Second Edition. New Jersey: Prentice-Hall Inc., 1970.

Deese, J. Structure of Association in Language and Thought. Baltimore : John Hopkins Press, 1965.

Ebel, Robert L. (ed). Encyclopedia of Educational Research. Fourth edition. London: The Macmillan Company, 1969.

Farr, R., et al. "An Examination of Reading Programs in Indiana Schools," Bulletin of School of Education, 45 (1972), 5-92.

Faucett, Lawrence, and Itsu Maki. A Study of English Word-Values Statistically Determined from the Latest Extensive Word Counts. Tokyo: Matsumura Sanshoda, 1932.

Ferguson, Charles A. "Introduction", The Learning of Language, ed. Carrol E. Reed, New York: Appleton-Century Crofts, 1971.

Fox, David J. The Research Process in Education. New York: Holt, Rinehart and Winston, Inc., 1969.

Francis, W. Nelson. Manual of Information to accompany Kucera, Henry, and W. Nelson Francis. Computational Analysis of Present-Day American English. Providence, Rhode Island: Brown University Press, 1967.

Fries, C. The Structure of English. New York: Harcourt, Brace and World, Inc., 1952.

Fries, Charles C., and A. Aileen Traver. English Word Lists: A Study of Their Adaptability for Instruction. Washington: American Council on Education, 1940.

Gates, Arthur I. A Reading Vocabulary for the Primary Grades. New York: Teachers College, Columbia University, 1926.

______ "Reflection and Return". Reading: A Human Right and A Human Problem, ed. Ralph C. Staiger and Oliver Andresen, Newark, Delaware: International Reading Association, 1968.

Glass, G.V., and J.C. Stanley. Statistical Methods in Education and Psychology. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1970.

Goodman, Kenneth S., et al. Choosing Materials to Teach Reading. Detroit: Wayne State University Press, 1966.

Gray, William S. The Teaching of Reading and Writing. U.N.E.S.C.O, 1969.

Gray, William S., and Bernice E. Leary. What Makes a Book Readable. Chicago: University of Chicago Press, 1935.

Harris, Albert J. How to Increase Reading Ability. Fifth edition. New York: David McKay Co. Inc., 1970.

Harris, Albert J., and Milton D. Jacobson. Basic Elementary Reading Vocabularies. New York: The Macmillan Company, 1972.

Harris, Chester W. (ed). Encyclopedia of Educational Research. Third edition. New York: Macmillan Co., 1960.

Horn, Ernest. Basic Writing Vocabulary. (Monographs in Education, First Series, No. 4), Iowa City: University of Iowa, 1926.

Huus, Helen. "Innovations in Reading Instruction: At Later Levels". The 67th Yearbook of the National Society for the Study of Education, Part II, ed. Helen M. Robinson, Chicago: The University of Chicago Press, 1968.

Jenkinson, M.D. "Information Gaps in Research in Reading Comprehension". Reading: Process and Pedagogy, ed. George B. Schick and Merrill M. May, Milwaukee: National Reading Conference, 1970, pp. 179-192.

Jewett, Arno (ed.). Improving Reading in the Junior High School. Washington, D.C: United States Government Printing Office, 1957.

Johnson, Marjorie Seddon. "Word Perception in the Reading - Thinking Process". Paper in Reading and Thinking, Proceedings of the 22nd Annual Reading Institute at Temple University, 1965, Temple University, Philadelphia, Pa. (ED 015 094)

Jones, Lyle V., and Joseph M. Wepman. A Spoken Word Count. Chicago, Illinois: Language Research Associates, 1966.

Jongsma, Eugene. The Cloze Procedure as a Teaching Technique. Newark, Delaware: The International Reading Association, 1971.

Klare, George R. The Measurement of Readability. Ames, Iowa: Iowa State University Press, 1963.

Kucera, Henry. "Computers in Language Analysis and Lexicography". The American Heritage Dictionary of the English Language, Boston: American Heritage Publishing Company, Inc., and Houghton Mifflin Company, 1969, p.xxxviii.

Kucera, Henry and W. Nelson Francis. Computational Analysis of Present-Day American English. Providence, Rhode Island: Brown University Press, 1967.

Luhn, H.P. "The Automatic Creation of Literature Abstracts". IBM Journal of Research and Development (April, 1958), 159-165. Reprinted in, Key Papers in Information Science, ed. Arthur W. Elias, Washington, D.C: American Society for Information Science, (1972), 87-93.

Malmquist, Eve. "Reading: A Human Right and A Human Problem". Reading: A Human Right and A Human Problem, ed. Ralph C. Staiger and Oliver Andressen, Newark, Delaware: International Reading Association, 1968.

Malsbary, Dean R. "A Study of the Terms that People Need to Understand in Order to Comprehend and Interpret the Business and Economic News Available Through the Mass Media", Studies in Education, Thesis Abstract Series, No. 4, Bloomington, Indiana: School of Education, Indiana University, 1952.

Maron, M.E. "Automatic Indexing: An Experimental Inquiry, Journal of the Association for Computing Machinery, (1961), 404-417. Reprinted in, Key Papers in Information Science, ed. Arthur W. Elias, Washington, D.C: American Society for Information Science, (1972), 94-107.

McLuhan, Marshall. Understanding Media: The Extensions of Man. New York: Signet Books, 1964.

Nason, H.M. "Efficient Reading - A Way to Permanent Education". Reading - The Right to Participate, Milwaukee, Wisconsin: Twentieth Yearbook of the National Reading Conference, 1971.

Paivio, A. Imagery and Verbal Processes. New York: Holt, Rinehart and Winston, Inc., 1971.

Pei, Mario A. The World's Chief Languages. Third edition. London: George Allen & Unwin Ltd., 1949.

Pratt, Edward. "Reading as a Thinking Process". Vistas in Reading, Volume II, Part I, ed. J. Allen Figurel, Seventh Annual Convention, International Reading Association, 1966.

Pressey, Luella C. Vocabulary Lists in Fifteen School Subjects. Bloomington, Illinois: Public School Publishing Co., 1924.

Rankin, E F. "Cloze Procedure - A Survey of Research", Fourteenth Yearbook of the National Reading Conference, eds, E.L.Thurston and L.E.Hafner, Milwaukee: National Reading Conference, 1965.

Rinsland, Henry D. A Basic Vocabulary of Elementary School Children. New York: The Macmillan Company, 1945.

Robinson, H.A. "Communications and Curriculum Change". Language, Reading, and the Communication Process, ed. Carl Braun, Newark, Delaware: International Reading Association Conference, 1971, p.2.

Rogers, John R. (ed). Linguistics in Reading Instruction. Mississippi: University of Mississippi, The Reading Clinic, 1965.

Russell, D.H. And H.R. Fea. "Research on Teaching Reading". Handbook of Research on Teaching, ed. N.L. Gage, Chicago: Rand McNally, 1963.

Singer, H., and R.B. Ruddell, (eds). Theoretical Models and Processes of Reading. Newark, Delaware: International Reading Association, 1970.

Smith, Nila Banton. American Reading Instruction. Newark, Delaware: International Reading Association, 1965.

Steinberg, S.H. Five Hundred Years of Printing. Bristol: Penguin Books, 1966.

Stothers, G.E., R.W.B. Jackson, and F.W. Minkler. A Canadian Word List. Toronto: The Ryerson Press, 1947.

Strang, Ruth., C.M. McCullough, and A.E. Traxler. The Improvement of Reading. New York: McGraw-Hill Book Company, 1967.

Tatsuoka, Maurice M., and David V. Tiedeman. "Statistics as an Aspect of Scientific Method in Research on Teaching", Handbook of Research on Teaching, ed. N.L. Gage, Chicago: Rand McNally, 1963.

Taylor, Stanford E., Helen Frackenpohl, and Catherine E. White. A Revised Core Vocabulary: A Basic Vocabulary for Grades 1-8: An Advanced Vocabulary for Grades 9-13. Research and Information Bulletin No. 5 (revised), Huntington, N.Y.: Educational Developmental Laboratories, 1949, 1969 (revised).

Thorndike, Edward L. The Teacher's Word Book. New York: Teachers College, Columbia University, 1921.

Thorndike, Edward L. A Teacher's Word Book of the Twenty Thousand Words Found Most Frequently and Widely in General Reading for Children and Young People. New York: Teachers College, Columbia University, 1931.

Thorndike, Edward L., and Irving Lorge. The Teacher's Word Book of 30,000 Words. New York: Teachers College, Columbia University, 1944.

Toffler, A. Future Shock. New York: Random House, 1970.

Traxler, Arthur E. "Development of a Vocabulary Test for High School Pupils and College Freshmen", 1962 Fall Testing Program in Independent Schools and Supplementary Studies, Educational Records Bulletin, No. 83 (February, 1963), 67-73.

Webb, Eugene J, et al. Unobtrusive Measures: Nonreactive Research in the Social Sciences. Chicago: Rand McNally & Co., 1966.

West, Michael. A Study of the Vocabulary of Children Before Entering First Grade. Washington: The International Kindergarten Union, 1928.

Yule, G. Udny. The Statistical Study of Literary Vocabulary. Cambridge, England, 1944.

B. PERIODICALS

Ames, Wilbur S. "The Development of a Classification Scheme of Contextual Aids", Reading Research Quarterly, 2 (Fall, 1966).

Aukerman, Robert C. "Readability of Secondary School Literature Textbooks: A First Report", English Journal, 54, (September, 1965).

Barton, Johnson D. "Computer Frequency Control of Vocabulary in Language Learning Reading Materials", Instructional Science, 1 (March, 1972), 121-131.

Beier, Ernst G., John A. Starkweather, and Dan E. Miller. "Analysis of Word Frequencies in Spoken Language of Children", Language and Speech, 10 (1967), 217-227.

Betts, Emmett A. "Structure in the Reading Program", Elementary English, XL (March, 1965), 238-242.

Bickley, A.C., et al. "The Cloze Procedure: A Conspectus", Journal of Reading Behavior, 2 (Summer, 1970), 232-243.

Bloomer, R.H. "Connotative Meaning and the Reading and Spelling Difficulty of Words", The Journal of Eudcational Research, 55 (November, 1961), 107-112.

Bormuth, John R. "Readability: A New Approach", Reading Research Quarterly, 1 (1966), 79-131.

Bormuth, John R. "New Developments in Readability Research", Elementary English, 44 (December, 1967), 840-845.

Bortnick, R., and G.S. Lopardo. "An Instructional Application of the Cloze Procedure", Journal of Reading, 16 (January, 1973), 296-299.

Broadbent, D.E. "Word Frequency Effect and Response Bias", Psychological Review, 74 (January, 1967), 1-10.

Bruner, J.S. "Neural Mechanisms in Perception", Psychological Review, 64 (1957), 340-358.

Card, William, and Virginia McDavid. "Frequencies of Structure Words in the Writing of Children and Adults", Elementary English, 42 (December, 1965), 878-882,894.

Chronister, G.M., and K.M. Ahrendt. "Reading Instruction in British Columbia's Secondary Schools", Journal of Reading, (March, 1968), 425-427.

Coleman, E.B. "Experimental Studies of Readability". Elementary English, XLV (March, 1968), 316-323,333.

Conway, James A., and Troy V. McKelvey. "The Role of the Relevant Literature: A Continuous Process", The Journal of Educational Research, 63 (May-June, 1970), 407-413.

Culhane, Joseph W. "Cloze Procedures and Comprehension", The Reading Teacher, 23 (February, 1970), 410-413.

Dale, Edgar., and Jeanne S. Chall. "A Formula for Predicting Readability", Educational Research Bulletin, 27 (1948), 11-20.

Dale, Edgar. "The Problem of Vocabulary in Reading", Educational Research Bulletin, 35 (1956), 113-123.

______ "VOCABULARY MEASUREMENT: TECHNIQUES AND MAJOR FINDINGS", Elementary English, 42 (December, 1965), 895-901.

Dodds, William J. "Highlights from the History of Reading Instruction", The Reading Teacher, 21 (December, 1967), 274-280.

Dolch, Edward W. "A Basic Sight Vocabulary", Elementary School Journal, 36 (1936), 456-460.

Durr, William K. "Computer Study of High Frequency Words in Popular Trade Juveniles", The Reading Teacher, 27, (October, 1973), 37-42.

Fry, Edward. "A Readability Formula That Saves Time", Journal of Reading, 11 (April, 1968), 513-516, 575-578.

Glazer, Susan Mandel. "Is Sentence Length a Valid Measure of Difficulty in Readability Formulas?", The Reading Teacher, 27 (February, 1974), 464-468.

Howes, Davis. "A Word Count of Spoken English", Journal of Verbal Learning and Verbal Behavior, 5 (1966), 572-604.

Johnson, Dale D. "The Dolch List Re-examined", The Reading Teacher, 24 (February, 1971), 449-457.

Johnson, D. Barton. Computer Frequency Control of Vocabulary in Language Learning Reading Materials", Instructional Science, 1, (March, 1972), 121-131.

Klare, George R. "Comments on Bormuth's Readability: A New Approach", Reading Research Quarterly, 4 (1966), 119-125.

Klare, George R. "The Role of Word Frequency in Readability Research", Elementary English, 45 (January, 1968), 12-22.

Kyte, George C. "A Core Vocabulary in the Language Arts", Phi Delta Kappan, 34 (March, 1953), 231-34.

Lively Bertha, and S.L Pressey. "A Method of Measuring the Vocabulary Burden of Textbooks", Educational Administration and Supervision, 9 (October, 1923), 389-398.

Lorge, Irving. "Word Lists as Background for Communication", Teachers College Record, 45 (May, 1944), 543-52.

Louthan, Vincent. "Some Systematic Grammatical Deletions and Their Effects on Reading Comprehension", English Journal, 54 (April, 1965), 295-299.

MacGinitie, Walter H., and Richard Tretiak. "Sentence Depth Measures as Predictors of Reading Difficulty", Reading Research Quarterly, VI (Spring, 1971), 364-377.

McLaughlin, Harry G. "SMOG Grading - a New Readability Formula", Journal of Reading, (May, 1969), 639-646.

Nyman, Patricia, et al. "An Attempt to Shorten the Word List with the Dale-Chall Readability Formula." Educational Research Bulletin, 40 (September, 1961), 150-152.

Palmer, William S. "Readability, Rhetoric, and the Reduction of Uncertainty", Journal of Reading , 19 (April, 1974), 552-558.

Patty, W.W., and W.I.Painter. "A Technique for Measuring the Vocabulary Burden of Textbooks", Journal of Educational Research, 24 (September, 1931), 127-134.

Powers, S.R. "The Vocabularies of High School Science Textbooks", Teachers College Record, 26 (January, 1925), 368-382.

Ramanauskas, Sigita. "The Responsiveness of Cloze Readability Measures to Linguistic Variables Operating Over Segments of Text Longer Than a Sentence", Reading Research Quarterly, 8 (Fall, 1972), 72-91.

Robinson, H. Alan and Dan S. Dramer. "High School Reading - 1958", Journal of Developmental Reading, 3 (Winter, 1960), 94-105. (See successive issues for summaries related to 1961 through 1966).

Robinson, Helen M., Samuel Weintraub, and Helen K. Smith. "Summary of Investigations Relating to Reading, July 1, 1966 to June 30, 1967", Reading Research Quarterly, 3 (Winter, 1967), 151-301. (See successive Winter issues for summaries related to 1968 through 1973).

Ross, Ramon Royal. "Frannie and Frank and the Flannelboard", The Reading Teacher, 27 (October, 1973), 43-47.

Siliakus, H.J. "Computer-Aided Word Research", Babel, 3 (July, 1967), 19-21.

Smith, John M., and Maxwell E. McCombs. "Research in Brief: The Graphics of Prose", Visible Language, V (Autumn, 1971) 365-369.

Smith, Nila Banton. "What Have We Accomplished in Reading? A Review of the Past Fifty Years". Elementary English, 38 (March, 1961), 141-150.

______ "The Many Faces of Reading Comprehension". The Reading Teacher, 23 (December, 1969)

Spache, George D. "A New Readability Formula for Primary -Grade Reading Materials", Elementary School Journal, 53 (March, 1953), 410-413.

Stauffer, R.G. "A Study of the Prefixes in the Thorndike List to Establish a List of Prefixes That Should be Taught in the Elementary School", Journal of Educational Research, 35, (1944), 453-458.

Stone, Clarence R. "Measuring Difficulty of Primary Reading Material: A Constructive Criticism of Spache's Measure", Elementary School Journal, 57 (October, 1956), 36-41.

Summers, Edward G. "Important Resource for Secondary Reading", Journal of Reading, 10 (November, 1966), 88-102.

Summers, E.G., Brother Leonard Courtney, and Peter Edwards. "Guide to Professional Textbooks and Research in Secondary Reading Instruction," The English Quarterly, 7 (Summer, 1974), 124-146.

Taylor, W. "Cloze Procedure: A New Tool for Measuring Readability", Journalism Quarterly, 30, 1953, 414-433.

Townsend, A. "Reading in the Junior Grades", The Reading Teacher, 15 (March, 1962), 369-371.

Vogel, Mabel, and W. Carleton Washbourne. "An Objective Method of Determining Grade Placement of Children's Reading Materials", Elementary School Journal, 28 (January, 1928), 373-381.

Warfel, Harry R. "A Bag With Holes", Journal of Developmental Reading, III (Autumn, 1959), 320-333.

Washbourne, Carleton W., and Mabel V. Morphett. "Grade Placement of Children's Books", Elementary School Journal, 38 (January, 1938), 355-364.

Weintraub, Samuel. "The Cloze Procedure", The Reading Teacher, 21 (March, 1968), 567,569,571,607.

Zipf, George Kingsley. "The Meaning-Frequency Relationship of Words", The Journal of General Psychology, 33 (October, 1945), 251-256.

UNPUBLISHED MATERIALS

Aaronson, Shirley. "Vocabulary Instruction: Challenge of the 70's". Paper read at the National Reading Conference, December, 1971, Tampa, Florida. (ED 058 016)

Artley, A. Sterl. Trends and Practices in Secondary School Reading: A Companion Bibliography to A. Sterl Artley's Monograph. Bloomington, Indiana: Indiana University, March, 1970. (AA 000 507)

Aulls, Mark W. "Toward a Systematic Approach to How the Reader Uses Context to Determine Meaning". Paper read at the National Reading Conference, December, 1970, St. Petersbury, Florida. (ED 049 003)

Austin, Warren B. A Computer-Aided Technique for Stylistic Discrimination: The Authorship of Greene's "Groatsworth of Wit". Final Report. 1969. (ED 030 322)

Bailey, Stephen D. "Recent Trends and Developments in Research Involving the Cloze Procedure". Unpublished research paper. Faculty of Education, University of British Columbia, 1973.

Berg, P. C. "The Psychology of Reading Behavior". Paper read at the National Reading Conference, December, 1968, Los Angeles. (ED 028 050)

Berkeley, Edmund C. Research in Computer-Assisted Explanation Applied to Navy Training Manuals. Final Report. Springfield, Virginia: National Technical Information Service, 1972. (ED 074 729)

Bormuth, John R. "New Data on Readability". Paper read at a meeting cosponsored by the International Reading Association and the American Educational Research Association, May, 1967, Seattle. (ED 016 586)

______ Cloze Readability Procedure. Report Number CSEIP-OR-1, Los Angeles: University of California, 1967. (ED 010 983)

______ "Empirical Determination of the Instructional Reading Level". Paper read at the International Reading Association Conference, April, 1968, Boston. (ED 020 084)

______ "The Effectiveness of Current Procedures for Teaching Reading Comprehension". Paper read at the Fifty-Eighth Annual Meeting of the National Council of Teachers of English, November, 1968, Milwaukee.

______ Development of Readability Analyses. Project No. 7-0052, U.S. Office of Education, March 1969.

______ "EDUPLAN Bibliography". Unpublished bibliography, University of Chicago, 1972.

Botel, M. "Ascertaining Instructional Levels". Paper read at the International Reading Association Conference, May, 1967, Seattle. (ED 014 373)

British Columbia Department of Education. Prescribed Textbooks, 1972-73. Grades I-XII. Victoria: Curriculum Development Branch, 1972.

Bruner, Jerome. "Well Begun is Half Done: Thoughts About Early Childhood". Address at the International Reading Association Annual Convention, May, 1972, Detroit, Michigan.

Carver, Ronald P. "What is Reading Comprehension and How Should it be Measured?". Paper read at the National Reading Conference, December, 1969, Atlanta, Georgia. (ED 038 243)

Carroll, John B. "Behind the Scenes in the Making of a Corpus-Based Dictionary and a Word Frequency Book". Paper read at the meeting of the National Council of Teachers of English, November, 1971, Las Vegas, Nevada. (ED 056 842)

Chall, J. "Research in Linguistics and Reading Instruction: Implications for Further Research and Practice". Paper read at the International Reading Association Conference, April, 1968, Boston, Mass. (ED 028 904)

Cooper, J. L. "The Reading Program Spans the Total Curriculum". Paper read at the International Reading Association Conference, May, 1967, Seattle. (ED 015 824)

Cronnell, Bruce. Designing a Reading Program Based on Research Findings in Orthography. 1971. (ED 057 990)

Dauzat, S. V. "Structure Word Usage in the Verbal Discourse of Two Groups of Children". Unpublished Doctor's dissertation, The University of Mississippi, 1968.

Dunn-Rankin, Peter. "Analyzing the Development of Reading Skill Using an Error-Word Preference Inventory". Paper read at the meeting of the American Educational Research Association, February, 1971, New York, N.Y. (ED 051 960)

Eagan, Sister Ruth Louise. "An Investigation into the Relationship of the Pausing Phenomena in Oral Reading and Reading Comprehension". Unpublished Doctor's dissertation, the University of Alberta, 1973.

Fagan, William T. "An Investigation into the Relationship Between Reading Difficulty and the Number of Types of Sentence Transformations". Unpublished Doctor's dissertation, the University of Alberta, 1969.

Frost, Joe L. "Application of Structure Process Theory to the Teaching of Reading". Paper read at the National Conference of Teachers of English, March, 1970, St Louis, Missouri. (ED 045 288)

Fuellhart, Patricia O., and David C. Weeks. Compilation and Analysis of Lexical Resources in Information Science. Final Report. Springfield, Virginia: Clearinghouse for Federal Scientific and Technical Information, 1968. (ED 021 602)

Geyer, James R. Cloze Procedure as a Predictor of Comprehension in Secondary Social Studies Materials. Olympia: Washington State Board for Community College Education, 1968. (ED 039 157)

Guthrie, J. T. Learnability Versus Readability of Texts. Baltimore, Maryland: Center for Social Organization of Schools, John Hopkins University, 1970. (ED 042 594)

Harris, Jessica L. A Study of the Computer Arrangeability of Complex Terms Occurring in a Major Tool Used in Subject Analysis. Final Report, 1969. (ED 028 793)

Hater, M A. "The Cloze Procedure as a Measure of the Reading Comprehensibility and Difficulty of Mathematical English." Unpublished Doctor's dissertation, Purdue University, 1969.

Hill, Walter R., and Norma G. Bartin. Secondary Reading Programs: Description and Research. ERIC/CRIER Reading Review Series, July 1971. (ED 055 759).

Hill, Walter R., and Norma G Bartin. Reading Programs in Secondary Schools: An Annotated Bibliography. Newark, Delaware: International Reading Association, 1971. (ED 071 057)

Houska, J.T. "The Efficiency of the Cloze Procedure as a Readability Tool on Technical Content Material Used in Industrial Education at the High School Level." Unpublished Doctor's dissertation, University of Illinois at Urbana and Champaign, 1971.

Jacobs, H. Donald. A Free-Association Word List for the Willamette Valley. Eugene: Oregon University, September, 1967. (ED 015 845)

Jacobson, Milton D. "Reading Difficulty of Physics and Chemistry Textbooks in Use in Minnesota." Unpublished Doctor's dissertation, University of Minnesota, 1961.

Jacobson, Milton D. "Developing and Comparing Elementary School Word Lists by Computer". Paper read at the meeting of the American Educational Research Association, April, 1972, Chicago. (ED 062 102)

Jacobson, Milton D., and Mary Ann MacDougall. "Computerized Model of Program Structure and Learning Difficulty". Paper in Proceedings of the 1969 Convention of the American Psychological Association, Washington, D.C. (ED 040 006)

Jongsma, Eugene R. The Cloze Procedure: A Survey of the Research. Indiana University: School of Education, 1970, (ED 050 893)

Kulm, G. "Measuring the Readability of Elementary Algebra Using the Cloze Technique." Paper presented at the Annual Meeting of the American Educational Research Association, February, 1971.

Lefevre, Carl A. "Language and Critical Reading: The Consummate Reader". Paper read at the National Reading Conference, December, 1969, Atlanta, Georgia. (ED 038 249)

Lerner, J. W. A Global Theory of Reading... and Linguistics. Newark, Delaware: International Reading Association, February, 1968. (ED 023 538)

Lott, Deborah, et al. _Functional Equivalence of Feature Combinations in the Visual Identification of Words._ Inglewood, California: Southwest Regional Educational Laboratory, 1968. (ED 035 516)

Lynch, Mervin D., et al. "The Building Block Construct as a Possible Model for Decoding Processes". Paper read at the National Reading Conference, December, 1970, St. Petersburg, Florida. (ED 049 002)

MacGinitie, W. H., and R. Tretiak. "Measures of Sentence Complexity as Predictors of the Difficulty of Reading Materials". In Proceedings of the 77th Annual Convention of the American Psychological Association, 1969. (ED 038 254)

Miller, Allan. _Programmer's Guide to the Edwards' Corpus._ Vancouver, British Columbia: Computing Centre, University of British Columbia, 1974.

Olsen H. C. "Linguistic Principles and the Selection of Materials". Paper read at the International Reading Association conference, April, 1968, Boston, Mass. (ED 022 649)

Potter, Thomas C. _A Taxonomy of Cloze Research, Part I: Readability and Reading Comprehension._ Inglewood, California: Southwest Regional Educational Laboratory, 1968. (ED 035 514)

Rawson, Hildred I. "A Study of the Relationships and Development of Reading and Cognition". Unpublished Doctor's dissertation, The University of Alberta, 1969.

Rosenshine, Barak. "New Correlates of Readability and Listenability." Paper read at the International Reading Association conference, April, 1968, Boston, Mass. (ED 024 528)

Seels, Barbara, and Dale Edgar. _Readability and Reading. An Annotated Bibliography: 1971 Revision._ Newark, Delaware: International Reading Association, 1971. (ED 049 896)

Shima, Fred. _Research on Word Association in Connected Discourse._ Inglewood, California: Southwest Regional Educational Laboratory, 1970. (ED 043 470)

Summers, Edward G. An Annotated Bibliography of Selected Research Related to Teaching Reading in the Secondary School: 1900-1960. University of Pittsburgh: School of Education, 1963. (ED 010 757).

______ An Annotated Bibliography of Selected Research Related to Teaching Reading in the Secondary School Supplement: 1961-1963. University of Pittsburgh: School of Education, 1964. (ED 010 758).

______ International Reading Association Conference Proceedings Reports on Secondary Reading. Bloomington, Indiana: ERIC/CRIER, 1967. (ED 013 185)

Summers, Edward G., Charles H. Davis, and Catherine F. Siffin. Published Research Literature in Reading: 1900-1949. ERIC Document Reproduction Service, Bethesda, Md., 1968. (ED 013 970).

______ Published Research Literature in Reading: 1950-1963. ERIC Document Reproduction Service, Bethesda, Md., 1967. (ED 012 834).

______ Published Research Literature in Reading: 1964-1966. ERIC Document Reproduction Service, Bethesda, Md., 1968. (ED 013 969).

Vernon, Evelyn, I. "Words Make For Success". Paper read at the International Reading Association, April-May, 1969, Kansas City, Missouri. (ED 034 662)

Weaver, Wendell W., and A. C. Bickley. "Structural-Lexical Predictability of Materials Which Predictor Has Previously Produced or Read". Paper in the 1967 Proceedings of the American Psychological Association, Division 15. (ED 011 812)

Whipple, Gertrude. "Practical Problems of Schoolbook Selection for Disadvantaged Pupils". Paper read at the International Reading Association Conference, April, 1968, Boston, Mass. (ED 029 750)

Wolfe, Josephine B. "Applying Research Findings in Comprehension to Classroom Practice". Paper read at the International Reading Association Conference, May, 1967, Seattle. (ED 014 371)

Young, Carol E. Development of Language Analysis Procedures With Application to Automatic Indexing. Columbus Ohio: Computer and Information Science Research Center, 1973, p.69. (ED 078 843)

APPENDIX A

INDEX OF TEXTS AND SAMPLES BY GRADE LEVEL

C. ENGLISH 8.

(Total of 17 Samples)

*1C01C Text: The Craft of Writing. Don Mills, Ontario: Longmans Canada Ltd., 1965.

Author: R.J. McMaster.

Sample	Pages	Sample	Pages
01	2-3	05	96-97
02	25-26	06	119-121
03	48-49	07	132-136
04	71-75		

*1C02C Text: Short Stories of Distinction. Agincourt: The Book Society of Canada Ltd., 1960.

Author: L.H. Newell and J.W. MacDonald (eds).

Sample	Pages	Sample	Pages
01	9-10	06	124-125
02	32-33	07	147-148
03	55-56	08	170-171
04	78-79	09	192-194
05	101-102	10	215-216

D. HOME ECONOMICS 8.

(Total of 22 Samples)

*1D01C Text: Teen Guide to Homemaking. Toronto: McGraw-Hill Co. of Canada Ltd., 1968.

Authors: M.S. Barclay and F. Champion.

Sample	Pages	Sample	Pages
01	6-8	12	230-235
02	34-36	13	247-248
03	52-53	14	262-265
04	71-72	15	278-280
05	85-86	16	306-308
06	108-111	17	334-336
07	124-125	18	342-345
08	153-154	19	366-368
09	168-170	20	392-395
10	180-181	21	406-408
11	218-221	22	428-430

E. INDUSTRIAL EDUCATION 8.

(Total of 9 Samples)

*1E01C Text: Exploring Industrial Education. New York: McGraw-Hill Co. of Canada Ltd., 1968.

Authors: Jes Laustrup, et al.

Sample	Pages	Sample	Pages
01	3-5	06	141-144
02	32-38	07	161-162
03	51-55	08	181-183
04	55-106	09	196-197
05	106-115		

F. MATHEMATICS 8.

(Total of 14 Samples)

*1F01C Text: Introduction to Mathematics. Reading, Mass: Addison-Wesley Publishing Co. Inc., 1962.

Author: C.F. Brumfiel, et al.

Sample	Pages	Sample	Pages
01	14-16	08	175-179
02	31-35	09	188-197
03	52-56	10	201-207
04	70-74	11	226-228
05	101-102	12	243-244
06	131-133	13	259-260
07	136-141	14	264-268

G. SCIENCE 8.

(Total of 20 Samples)

*1G01C Text: Labtext in Science: Book 1. Toronto: The Copp Clark Publishing Co., 1968.

Authors: G.H. Cannon, et al.

Sample	Pages	Sample	Pages
01	13-16	06	121-124
02	25-27	07	138-140
03	61-62	08	163-164
04	78-80	09	179-180
05	100-101		

*1G02C Text: <u>Reading About Science 1</u>.
Holt, Rinehart & Winston of Canada, Ltd., 1968.

Authors: Clifford J. Anastasiou, et al.

Sample	Pages	Sample	Pages
01	12-13	07	143-144
02	43	08	157-158
03	57-58	09	176-177
04	72-73	10	200-201
05	90	11	226-227
06	122-123		

H. SOCIAL STUDIES 8.

(Total of 22 Samples)

*1H01C Text: <u>Man in the Tropics</u>. Scarborough, Ontario.
Bellhaven House Ltd., 1968.

Authors: Bordon E. Carswell, et al.

Sample	Pages	Sample	Pages
01	1-3	09	203-205
02	25-29	10	227-233
03	51-54	11	245-247
04	76-78	12	269-272
05	100-104	13	294-297
06	126-127	14	319-322
07	153-156	15	345-346
08	177-181		

*1H02C Text: <u>The Shaping of Modern Europe</u>. Toronto:
The MacMillan Company of Canada Ltd., 1968.

Author: Geoffrey Williams.

Sample	Pages	Sample	Pages
01	5-6	05	99-100
02	28-29	06	117-118
03	50-51	07	140-141
04	76-77		

B. COMMERCE	9.

(Total of 25 Samples)

*2B01C Text: Personal Typewriting. Toronto: W.J. Gage Ltd., 1967.

Authors: S.J. Wanous, et al.

Sample	Pages	Sample	Pages
01	preface	07	151-156
02	vi-vii	08	168
03	54-61	09	189-195
04	65-69	10	211-212
05	95-99	11	239-242
06	132-133		

*2B02C Text: The Junior Clerk. Toronto: Sir Isaac Pitman (Canada) Ltd., 1970.

Authors: C.A. Trotter and P.C. Glover.

Sample	Pages	Sample	Pages
01	1-3	08	171-174
02	32-35	09	187-189
03	60-61	10	204-213
04	75-77	11	238-239
05	95-97	12	261-263
06	125-126	13	279-280
07	133-148	14	295-298

C. ENGLISH 9.

(Total of 47 Samples)

*2C01C Text: Learning English. Toronto: The MacMillan Company of Canada Ltd., 1963.

Author: Philip G. Penner and Ruth E. McConnell.

Sample	Pages	Sample	Pages
01	1-3	11	230-234
02	24-26	12	256-263
03	49-50	13	284-288
04	55-57	14	310-314
05	70-73	15	337-340
06	95-101	16	360-363
07	123-124	17	384-386
08	146-158	18	411-412
09	181-183	19	435-441
10	202-208	20	453-455

*2C02C Text: The Accomplished Reader. Don Mills, Ontario: Bellhaven House, 1964.

Author: Maurice Gibbons and Alan Dawe.

Sample	Pages	Sample	Pages
01	1-2	05	96-97
02	26-27	06	118
03	50-51	07	142-143
04	73-74		

*2C03C Text: Prose Readings. Ontario: Longmans Canada Ltd., 1964.

Author: Jan de Bruyn (ed).

Sample	Pages	Sample	Pages
01	3-4	06	119-120
02	26-28	07	142-144
03	50-51	08	166-167
04	73-74	09	189-190
05	96-97	10	212-213

*2C04C Text: The Harrap Book of Modern Short Stories. Toronto: Clarke, Irwin & Co. Ltd., 1964.

Author: J.G. Bullocke (ed).

Sample	Pages	Sample	Pages
01	9-10	06	113-115
02	21-22	07	137-138
03	44-45	08	159-161
04	67-68	09	183-184
05	89-91	10	200-202

| D. HOME ECONOMICS 9. |

(Total of 76 Samples)

*2D01C Text: Guide to Modern Meals. Toronto: McGraw-Hill Co. of Canada Ltd., 1970.

Authors: D.E. Shank, et al.

Sample	Pages	Sample	Pages
01	2-3	12	246-249
02	31-32	13	267
03	61-61	14	289-291
04	72-73	15	306-307
05	97-98	16	324
06	120-121	17	342-344
07	137-141	18	365-366
08	159-163	19	383-384
09	186-187	20	417
10	206-207	20	417
11	223	21	426-427

*2D02C Text: Clothes for Teens. Toronto:
D.C. Heath Canada Ltd., 1970.

Authors: E. Todd and F. Roberts.

Sample	Pages	Sample	Pages
01	2-3	12	258
02	35	13	283-284
03	62	14	299
04	81	15	328-329
05	108	16	338-339
06	125	17	359-361
07	147-148	18	376-378
08	167-168	19	400-401
09	194-196	20	439-440
10	214-215	21	460-461
11	240	22	489-490

*2D03C Text: Learning About Children. Philadelphia:
J.B. Lippincott Co., 1964.

Authors: R.M. Shuey, et al.

Sample	Pages	Sample	Pages
01	18-20	08	170-171
02	36-39	09	193-194
03	63-64	10	216-219
04	82-83	11	237-240
05	95-97	12	258-259
06	126-128	13	279-281
07	146-148	14	289-290

*2D04C Text: Up the Years From 1 to 6. Dept. of National Health & Welfare, Ottawa, Canada, 1967.

Author: Not given.

Sample	Pages	Sample	Pages
01	8-10	06	115-116
02	31-33	07	136-139
03	55-56	08	151-152
04	68-69	09	183-184
05	95-96	10	201-202

*2D05C Text: So - You Are Ready To Cook. Minneapolis: Burgess Publishing Co., 1964.

Author: M.A.Duffie.

Sample	Pages	Sample	Pages
01	4-6	06	127-130
02	37	07	141-143
03	59-61	08	162
04	86-87	09	182-183
05	103		

E. INDUSTRIAL EDUCATION 9.

(Total of 54 Samples)

*2E01C Text: General Woodworking. Toronto: McGraw-Hill Co. of Canada Ltd., 1965.

Author: Chris. H. Groneman.

Sample	Pages	Sample	Pages
01	1-2	08	175
02	42-44	09	184-186
03	54-55	10	210-211
04	74-76	11	235-238
05	95-113	12	253-254
06	114-135	13	272-273
07	158-160		

*2E02C Text: General Metals. Toronto: McGraw-Hill Co. of Canada Ltd., 1965.

Author: John L. Feirer.

Sample	Pages	Sample	Pages
01	12-14	09	185-187
02	35-36	10	210-212
03	59-61	11	226-227
04	67-69	12	238-240
05	104-105	13	260-262
06	126-129	14	273-274
07	149-151	15	317-319
08	167-170	16	340-349

*2E03C Text: General Power Mechanics. Toronto: McGraw-Hill Co. of Canada Ltd., 1970.

Authors: Robert M. Worthington, et al.

Sample	Pages	Sample	Pages
01	18-21	14	297-298
02	35-37	15	328-331
03	57-59	16	342-348
04	79-81	17	367-369
05	100-101	18	390-391
06	127-129	19	413-414
07	148-149	20	435
08	169-170	21	458-460
09	192-193	22	473-474
10	210-212	23	500-501
11	237-240	24	521-522
12	260-261	25	546-547
13	281-283		

F. MATHEMATICS 9.

(Total of 7 Samples)

*2F01C Text: Modern General Mathematics. Don Mills, Ontario: Addison-Wesley (Canada) Ltd., 1966.

Authors: R.E. Eicholy, et al.

Sample	Pages	Sample	Pages
01	1-70	05	146-161
02	75-106	06	183-214
03	108-127	07	227-331
04	130-143		

G. SCIENCE 9.

(Total of 24 Samples)

*2G01C Text: Developing Science Concepts in the Laboratory. Scarborough, Ontario: Prentice-Hall of Canada Ltd., 1968.

Authors: W.H. Rasmussen and M.C. Schmid.

Sample	Pages	Sample	Pages
01	1-2	08	156-159
02	22-23	09	171-175
03	41-44	10	199-202
04	76-77	11	222-224
05	92-94	12	248-251
06	111-116	13	278
07	136-138		

*2G02C Text: Reading About Science 2.
Holt, Rinehart & Winston of Canada Ltd., 1969.

Authors: M. Forster, et al.

Sample	Pages	Sample	Pages
01	13-14	07	145-146
02	30	08	172-173
03	63-64	09	191-192
04	78-79	10	218
05	100-101	11	241-242
06	124-125		

H. SOCIAL STUDIES 9.

(Total of 13 Samples)

*2H01C Text: Man In The Great Community. Scarborough, Ontario:
Bellhaven House Ltd., 1969.

Authors: G.E. Carswell, et al.

Sample	Pages	Sample	Pages
01	1-3	05	95-100
02	22-27	06	111-114
03	61-62	07	132-134
04	82-85	08	155-156

*2H02C Text: Our World of Change. Toronto:
McGraw-Hill Company of Canada, Ltd., 1969.

Author: Hugh R. Innis.

Sample	Pages	Sample	Pages
01	13-14	04	71-72
02	20-21	05	104-106
03	49-51		

A. AGRICULTURE 10.

(Pilot Study: Not used in Corpus)
(Total of 21 Samples)

*3A01C Text: Farmer's Shop Book. Milwaukee: The Bruce Publishing Co., 1953.

Authors: L.M. Roehl and A.D. Longhouse.

Sample	Pages	Sample	Pages
01	14-17	12	231-233
02	34-36	13	261
03	62-64	14	274
04	78-81	15	313-316
05	109-111	16	328-329
06	129	17	353-355
07	146-147	18	375
08	161-164	19	390-391
09	189-190	20	416-417
10	216-220	21	433-437
11	223-275		

B. COMMERCE 10.

(Total of 16 Samples)

*3B01C Text: New Basic Course in Pitman Shorthand. Toronto: Sir Isaac Pitman (Canada) Ltd., 1964.

Author: Not given.

Sample	Pages	Sample	Pages
01	viii-ix	05	88-89
02	27-37	06	113-121
03	55-66	07	137-145
04	77-87		

*3B02C Text: Exploring Business. Toronto: McGraw-Hill Co., of Canada Ltd., 1968.

Authors: J. Frank Dame, et al.

Sample	Pages	Sample	Pages
01	5-7	06	114-117
02	27-29	07	143
03	54-56	08	178-179
04	85	09	186
05	100-102		

C. ENGLISH 10.

(Total of 16 Samples)

*3C01C Text: Eighteen Stories. Don Mills, Ontario: J.M. Dent & Sons (Canada), 1965.

Authors: Malcolm Ross and John Stevens (eds).

Sample	Pages	Sample	Pages
01	1-3	07	140-141
02	25-26	08	163-165
03	48-49	09	187-188
04	71-72	10	210-211
05	94-95	11	233-234
06	117-118	12	251-252

*3C02C Text: Drama IV. Toronto: The MacMillan Co. of Canada Ltd., 1965.

Author: Herman Voaden (ed).

Sample	Pages	Sample	Pages
01	2-3	03	226-227
02	142-143	04	383-384

F. MATHEMATICS 10.

(Total of 14 Samples)

*3F01C Text: Mathematics: A Modern Approach.
Don Mills, Ontario:
Addison Wesley (Canada) Ltd., 1966.

Authors: M.S. Wilcox and J.E. Yarnelle.

Sample	Pages	Sample	Pages
01	1-5	08	190-191
02	14-20	09	207-209
03	54-57	10	237-263
04	65-66	11	295-297
05	90-91	12	304-305
06	101-102	13	322-324
07	150-153	14	346-347

G. SCIENCE 10.

(Total of 31 Samples)

*3G01C Text: Extending Science Concepts in the Laboratory.
Scarborough, Ontario:
Prentice-Hall of Canada Ltd., 1970.

Author: M.C. Schmid (ed).

Sample	Pages	Sample	Pages
01	1-5	10	204-210
02	31-37	11	253-256
03	55-56	12	259-264
04	79-85	13	291-298
05	106-108	14	311-319
06	126-128	15	323-326
07	149-154	16	329-336
08	164-169	17	372-375
09	193-200		

*3G02C Text: Reading About Science 3. Toronto: Holt, Rinehart & Winston of Canada, Ltd., 1970.

Author: J. Woodrow.

Sample	Pages	Sample	Pages
01	38-42	08	186-188
02	59	09	203-204
03	75	10	227-228
04	100-101	11	234-235
05	128-129	12	254-255
06	139-143	13	277-278
07	155	14	301-302

H. SOCIAL STUDIES 10.

(Total of 42 Samples)

*3H01C Text: A Regional Geography of North America. Toronto: Gage Educational Pub. Ltd., 1970.

Authors: G.S. Tomkins, et al.

Sample	Pages	Sample	Pages
01	11	16	320
02	26-27	17	327-328
03	53	18	347
04	74	19	372
05	97	20	387
06	112	21	406
07	134-135	22	434
08	159	23	449
09	175	24	469
10	196	25	490
11	210-211	26	514-515
12	236-237	27	536
13	254	28	557-559
14	279	29	581-582
15	299	30	596

*3H02C Text: <u>A Nation Developing: A Brief History of Canada</u>. Toronto: McGraw-Hill Company of Canada Ltd., 1970.

Author: J.A. Lower.

Sample	Pages	Sample	Pages
01	14-15	07	131-132
02	35-36	08	158-159
03	54-55	09	178-179
04	77-78	10	197-198
05	92-93	11	213-214
06	115-116	12	230-231

APPENDIX B

SAMPLE SIZES IN ALPHABETICAL ORDER AND ASCENDING RANK

SAMPLES IN ALPHABETICAL ORDER

SAMPLE	SIZE	SAMPLE	SIZE
*1C01C01	523	*1E01C02	521
*1C01C02	507	*1E01C03	612
*1C01C03	496	*1E01C04	579
*1C01C04	455	*1E01C05	556
*1C01C05	485	*1E01C06	505
*1C01C06	508	*1E01C07	473
*1C01C07	526	*1E01C08	473
*1C02C01	470	*1E01C09	441
*1C02C02	475	*1F01C01	453
*1C02C03	484	*1F01C02	459
*1C02C04	575	*1F01C03	498
*1C02C05	515	*1F01C04	427
*1C02C06	588	*1F01C05	481
*1C02C07	450	*1F01C06	566
*1C02C08	529	*1F01C07	549
*1C02C09	526	*1F01C08	552
*1C02C10	493	*1F01C09	545
*1D01C01	505	*1F01C10	491
*1D01C02	387	*1F01C11	541
*1D01C03	491	*1F01C12	522
*1D01C04	576	*1F01C13	481
*1D01C05	384	*1F01C14	509
*1D01C06	557	*1G01C01	505
*1D01C07	618	*1G01C02	515
*1D01C08	584	*1G01C03	514
*1D01C09	577	*1G01C04	482
*1D01C10	554	*1G01C05	512
*1D01C11	573	*1G01C06	453
*1D01C12	480	*1G01C07	458
*1D01C13	509	*1G01C08	468
*1D01C14	560	*1G01C09	495
*1D01C15	427	*1G02C01	513
*1D01C16	391	*1G02C02	490
*1D01C17	535	*1G02C03	524
*1D01C18	635	*1G02C04	485
*1D01C19	436	*1G02C05	445
*1D01C20	466	*1G02C06	470
*1D01C21	611	*1G02C07	499
*1D01C22	571	*1G02C08	533
*1E01C01	464	*1G02C09	495

SAMPLES IN ALPHABETICAL ORDER

SAMPLE	SIZE	SAMPLE	SIZE
*1G02C10	526	*2B02C06	498
*1G02C11	525	*2B02C07	485
*1H01C01	532	*2B02C08	471
*1H01C02	522	*2B02C09	463
*1H01C03	543	*2B02C10	495
*1H01C04	512	*2B02C11	523
*1H01C05	495	*2B02C12	470
*1H01C06	562	*2B02C13	490
*1H01C07	500	*2B02C14	515
*1H01C08	537	*2C01C01	500
*1H01C09	512	*2C01C02	487
*1H01C10	494	*2C01C03	499
*1H01C11	512	*2C01C04	528
*1H01C12	493	*2C01C05	500
*1H01C13	519	*2C01C06	465
*1H01C14	499	*2C01C07	455
*1H01C15	496	*2C01C08	445
*1H02C01	495	*2C01C09	453
*1H02C02	508	*2C01C10	480
*1H02C03	479	*2C01C11	445
*1H02C04	501	*2C01C12	500
*1H02C05	490	*2C01C13	476
*1H02C06	527	*2C01C14	504
*1H02C07	477	*2C01C15	547
*2B01C01	451	*2C01C16	444
*2B01C02	469	*2C01C17	491
*2B01C03	470	*2C01C18	509
*2B01C04	361	*2C01C19	381
*2B01C05	618	*2C01C20	537
*2B01C06	498	*2C02C01	498
*2B01C07	530	*2C02C02	520
*2B01C08	458	*2C02C03	458
*2B01C09	551	*2C02C04	494
*2B01C10	608	*2C02C05	489
*2B01C11	480	*2C02C06	420
*2B02C01	571	*2C02C07	521
*2B02C02	528	*2C03C01	491
*2B02C03	545	*2C03C02	439
*2B02C04	489	*2C03C03	562
*2B02C05	448	*2C03C04	499

SAMPLES IN ALPHABETICAL ORDER

SAMPLE	SIZE	SAMPLE	SIZE
*2C03C05	530	*2D02C04	549
*2C03C06	470	*2D02C05	458
*2C03C07	474	*2D02C06	471
*2C03C08	572	*2D02C07	470
*2C03C09	483	*2D02C08	496
*2C03C10	515	*2D02C09	478
*2C04C01	532	*2D02C10	526
*2C04C02	451	*2D02C11	485
*2C04C03	514	*2D02C12	480
*2C04C04	514	*2D02C13	489
*2C04C05	503	*2D02C14	511
*2C04C06	513	*2D02C15	438
*2C04C07	500	*2D02C16	485
*2C04C08	508	*2D02C17	477
*2C04C09	505	*2D02C18	514
*2C04C10	502	*2D02C19	407
*2D01C01	534	*2D02C20	524
*2D01C02	479	*2D02C21	454
*2D01C03	525	*2D02C22	501
*2D01C04	452	*2D03C01	544
*2D01C05	487	*2D03C02	496
*2D01C06	446	*2D03C03	508
*2D01C07	455	*2D03C04	507
*2D01C08	444	*2D03C05	516
*2D01C09	508	*2D03C06	506
*2D01C10	457	*2D03C07	516
*2D01C11	507	*2D03C08	502
*2D01C12	479	*2D03C09	493
*2D01C13	473	*2D03C10	473
*2D01C14	520	*2D03C11	435
*2D01C15	469	*2D03C12	496
*2D01C16	508	*2D03C13	448
*2D01C17	515	*2D03C14	488
*2D01C18	517	*2D04C01	564
*2D01C19	504	*2D04C02	615
*2D01C20	465	*2D04C03	501
*2D01C21	454	*2D04C04	572
*2D02C01	513	*2D04C05	588
*2D02C02	525	*2D04C06	469
*2D02C03	504	*2D04C07	489

SAMPLES IN ALPHABETICAL ORDER

SAMPLE	SIZE	SAMPLE	SIZE
*2D04C08	524	*2E02C16	494
*2D04C09	488	*2E03C01	511
*2D04C10	522	*2E03C02	519
*2D05C01	518	*2E03C03	467
*2D05C02	539	*2E03C04	558
*2D05C03	524	*2E03C05	525
*2D05C04	475	*2E03C06	525
*2D05C05	486	*2E03C07	551
*2D05C06	506	*2E03C08	561
*2D05C07	504	*2E03C09	514
*2D05C08	491	*2E03C10	458
*2D05C09	556	*2E03C11	479
*2E01C01	535	*2E03C12	538
*2E01C02	511	*2E03C13	479
*2E01C03	448	*2E03C14	488
*2E01C04	455	*2E03C15	488
*2E01C05	657	*2E03C16	548
*2E01C06	536	*2E03C17	506
*2E01C07	400	*2E03C18	523
*2E01C08	446	*2E03C19	521
*2E01C09	486	*2E03C20	519
*2E01C10	445	*2E03C21	567
*2E01C11	338	*2E03C22	474
*2E01C12	404	*2E03C23	506
*2E01C13	414	*2E03C24	447
*2E02C01	503	*2E03C25	517
*2E02C02	508	*2F01C01	505
*2E02C03	457	*2F01C02	503
*2E02C04	504	*2F01C03	480
*2E02C05	476	*2F01C04	485
*2E02C06	502	*2F01C05	561
*2E02C07	356	*2F01C06	501
*2E02C08	508	*2F01C07	581
*2E02C09	573	*2G01C01	507
*2E02C10	484	*2G01C02	500
*2E02C11	472	*2G01C03	517
*2E02C12	528	*2G01C04	543
*2E02C13	490	*2G01C05	502
*2E02C14	467	*2G01C06	494
*2E02C15	471	*2G01C07	508

SAMPLES IN ALPHABETICAL ORDER

SAMPLE	SIZE	SAMPLE	SIZE
*2G01C08	540	*3B02C04	477
*2G01C09	512	*3B02C05	435
*2G01C10	523	*3B02C06	499
*2G01C11	572	*3B02C07	429
*2G01C12	511	*3B02C08	495
*2G01C13	519	*3B02C09	458
*2G02C01	496	*3C01C01	509
*2G02C02	516	*3C01C02	570
*2G02C03	452	*3C01C03	523
*2G02C04	513	*3C01C04	551
*2G02C05	484	*3C01C05	564
*2G02C06	522	*3C01C06	519
*2G02C07	514	*3C01C07	594
*2G02C08	511	*3C01C08	612
*2G02C09	524	*3C01C09	532
*2G02C10	460	*3C01C10	620
*2G02C11	538	*3C01C11	538
*2H01C01	532	*3C01C12	556
*2H01C02	544	*3C02C01	454
*2H01C03	638	*3C02C02	522
*2H01C04	559	*3C02C03	461
*2H01C05	557	*3C02C04	430
*2H01C06	508	*3F01C01	499
*2H01C07	513	*3F01C02	450
*2H01C08	557	*3F01C03	508
*2H02C01	507	*3F01C04	484
*2H02C02	472	*3F01C05	450
*2H02C03	525	*3F01C06	505
*2H02C04	497	*3F01C07	549
*2H02C05	546	*3F01C08	531
*3B01C01	543	*3F01C09	478
*3B01C02	552	*3F01C10	492
*3B01C03	483	*3F01C11	537
*3B01C04	494	*3F01C12	541
*3B01C05	573	*3F01C13	546
*3B01C06	419	*3F01C14	530
*3B01C07	482	*3G01C01	548
*3B02C01	482	*3G01C02	477
*3B02C02	403	*3G01C03	517
*3B02C03	428	*3G01C04	497

SAMPLES IN ALPHABETICAL ORDER

SAMPLE	SIZE	SAMPLE	SIZE
*3G01C05	486	*3H01C14	437
*3G01C06	520	*3H01C15	489
*3G01C07	470	*3H01C16	503
*3G01C08	517	*3H01C17	478
*3G01C09	458	*3H01C18	482
*3G01C10	532	*3H01C19	428
*3G01C11	517	*3H01C20	431
*3G01C12	514	*3H01C21	455
*3G01C13	543	*3H01C22	520
*3G01C14	492	*3H01C23	439
*3G01C15	498	*3H01C24	494
*3G01C16	510	*3H01C25	469
*3G01C17	498	*3H01C26	466
*3G02C01	523	*3H01C27	503
*3G02C02	451	*3H01C28	530
*3G02C03	538	*3H01C29	480
*3G02C04	497	*3H01C30	377
*3G02C05	495	*3H02C01	460
*3G02C06	435	*3H02C02	449
*3G02C07	522	*3H02C03	492
*3G02C08	493	*3H02C04	442
*3G02C09	511	*3H02C05	490
*3G02C10	469	*3H02C06	423
*3G02C11	467	*3H02C07	470
*3G02C12	513	*3H02C08	447
*3G02C13	522	*3H02C09	537
*3G02C14	555	*3H02C10	485
*3H01C01	482	*3H02C11	541
*3H01C02	568	*3H02C12	456
*3H01C03	514		
*3H01C04	480		
*3H01C05	525		
*3H01C06	565		
*3H01C07	501		
*3H01C08	535		
*3H01C09	552		
*3H01C10	497		
*3H01C11	567		
*3H01C12	423		
*3H01C13	546		

SAMPLES RANKED IN ASCENDING ORDER

SAMPLE	SIZE	SAMPLE	SIZE
*2E01C11	338	*2E01C08	446
*2E02C07	356	*2D01C06	446
*2B01C04	361	*3H02C08	447
*3H01C30	377	*2E03C24	447
*2C01C19	381	*2B02C05	448
*1D01C05	384	*2D03C13	448
*1D01C02	387	*2E01C03	448
*1D01C16	391	*3H02C02	449
*2E01C07	400	*1C02C07	450
*3B02C02	403	*3F01C05	450
*2E01C12	404	*3F01C02	450
*2D02C19	407	*2C04C02	451
*2E01C13	414	*2B01C01	451
*3B01C06	419	*3G02C02	451
*2C02C06	420	*2D01C04	452
*3H02C06	423	*2G02C03	452
*3H01C12	423	*2C01C09	453
*1D01C15	427	*1F01C01	453
*1F01C04	427	*1G01C06	453
*3B02C03	428	*2D01C21	454
*3H01C19	428	*3C02C01	454
*3B02C07	429	*2D02C21	454
*3C02C04	430	*3H01C21	455
*3H01C20	431	*1C01C04	455
*2D03C11	435	*2E01C04	455
*3G02C06	435	*2D01C07	455
*3B02C05	435	*2C01C07	455
*1D01C19	436	*3H02C12	456
*3H01C14	437	*2D01C10	457
*2D02C15	438	*2E02C03	457
*3H01C23	439	*2D02C05	458
*2C03C02	439	*2B01C08	458
*1E01C09	441	*3B02C09	458
*3H02C04	442	*3G01C09	458
*2C01C16	444	*2E03C10	458
*2D01C08	444	*1G01C07	458
*1G02C05	445	*2C02C03	458
*2E01C10	445	*1F01C02	459
*2C01C08	445	*3H02C01	460
*2C01C11	445	*2G02C10	460

SAMPLES RANKED IN ASCENDING ORDER

SAMPLE	SIZE	SAMPLE	SIZE
*3C02C03	461	*2D02C17	477
*2B02C09	463	*3G01C02	477
*1E01C01	464	*3B02C04	477
*2C01C06	465	*3H01C17	478
*2D01C20	465	*2D02C09	478
*3H01C26	466	*3F01C09	478
*1D01C20	466	*2D01C02	479
*2E02C14	467	*1H02C03	479
*2E03C03	467	*2E03C11	479
*3G02C11	467	*2D01C12	479
*1G01C08	468	*2E03C13	479
*2D04C06	469	*2F01C03	480
*2B01C02	469	*2C01C10	480
*2D01C15	469	*2B01C11	480
*3G02C10	469	*2D02C12	480
*3H01C25	469	*3H01C04	480
*2C03C06	470	*1D01C12	480
*1G02C06	470	*3H01C29	480
*2D02C07	470	*1F01C13	481
*2B01C03	470	*1F01C05	481
*2B02C12	470	*3B01C07	482
*3G01C07	470	*3H01C01	482
*1C02C01	470	*3B02C01	482
*3H02C07	470	*1G01C04	482
*2D02C06	471	*3H01C18	482
*2E02C15	471	*2C03C09	483
*2B02C08	471	*3B01C03	483
*2H02C02	472	*2E02C10	484
*2E02C11	472	*1C02C03	484
*1E01C07	473	*2G02C05	484
*2D03C10	473	*3F01C04	484
*1E01C08	473	*2D02C11	485
*2D01C13	473	*3H02C10	485
*2C03C07	474	*1C01C05	485
*2E03C22	474	*1G02C04	485
*2D05C04	475	*2D02C16	485
*1C02C02	475	*2F01C04	485
*2E02C05	476	*2B02C07	485
*2C01C13	476	*3G01C05	486
*1H02C07	477	*2E01C09	486

SAMPLES RANKED IN ASCENDING ORDER

SAMPLE	SIZE	SAMPLE	SIZE
*2D05C05	486	*3G02C05	495
*2C01C02	487	*1G02C09	495
*2D01C05	487	*2D02C08	496
*2D04C09	488	*1H01C15	496
*2E03C14	488	*1C01C03	496
*2D03C14	488	*2G02C01	496
*2E03C15	488	*2D03C02	496
*3H01C15	489	*2D03C12	496
*2B02C04	489	*3H01C10	497
*2D04C07	489	*3G01C04	497
*2C02C05	489	*2H02C04	497
*2D02C13	489	*3G02C04	497
*2E02C13	490	*2C02C01	498
*2B02C13	490	*2B01C06	498
*3H02C05	490	*2B02C06	498
*1G02C02	490	*1F01C03	498
*1H02C05	490	*3G01C17	498
*1D01C03	491	*3G01C15	498
*1F01C10	491	*2C03C04	499
*2C03C01	491	*1H01C14	499
*2C01C17	491	*2C01C03	499
*2D05C08	491	*3F01C01	499
*3H02C03	492	*3B02C06	499
*3G01C14	492	*1G02C07	499
*3F01C10	492	*2C04C07	500
*2D03C09	493	*1H01C07	500
*1H01C12	493	*2C01C12	500
*1C02C10	493	*2G01C02	500
*3G02C08	493	*2C01C01	500
*3B01C04	494	*2C01C05	500
*3H01C24	494	*3H01C07	501
*2C02C04	494	*2D04C03	501
*2E02C16	494	*2F01C06	501
*1H01C10	494	*2D02C22	501
*2G01C06	494	*1H02C04	501
*2B02C10	495	*2C04C10	502
*1H02C01	495	*2G01C05	502
*3B02C08	495	*2D03C08	502
*1G01C09	495	*2E02C06	502
*1H01C05	495	*2C04C05	503

SAMPLES RANKED IN ASCENDING ORDER

SAMPLE	SIZE	SAMPLE	SIZE
*3H01C16	503	*2D02C14	511
*2E02C01	503	*2G02C08	511
*2F01C02	503	*2E03C01	511
*3H01C27	503	*3G02C09	511
*2C01C14	504	*2E01C02	511
*2E02C04	504	*2G01C12	511
*2D02C03	504	*1H01C11	512
*2D05C07	504	*1G01C05	512
*2D01C19	504	*1H01C09	512
*1G01C01	505	*1H01C04	512
*2C04C09	505	*2G01C09	512
*3F01C06	505	*2C04C06	513
*1E01C06	505	*2H01C07	513
*2F01C01	505	*2D02C01	513
*1D01C01	505	*2G02C04	513
*2D05C06	506	*3G02C12	513
*2E03C17	506	*1G02C01	513
*2D03C06	506	*2C04C03	514
*2E03C23	506	*3H01C03	514
*2H02C01	507	*2E03C09	514
*1C01C02	507	*1G01C03	514
*2G01C01	507	*2C04C04	514
*2D01C11	507	*3G01C12	514
*2D03C04	507	*2D02C18	514
*2H01C06	508	*2G02C07	514
*2C04C08	508	*2B02C14	515
*2E02C08	508	*1G01C02	515
*2E02C02	508	*2D01C17	515
*1H02C02	508	*2C03C10	515
*1C01C06	508	*1C02C05	515
*2G01C07	508	*2G02C02	516
*2D01C16	508	*2D03C07	516
*3F01C03	508	*2D03C05	516
*2D01C09	508	*3G01C03	517
*2D03C03	508	*2E03C25	517
*1F01C14	509	*2G01C03	517
*2C01C18	509	*2D01C18	517
*1D01C13	509	*3G01C08	517
*3C01C01	509	*3G01C11	517
*3G01C16	510	*2D05C01	518

SAMPLES RANKED IN ASCENDING ORDER

SAMPLE	SIZE	SAMPLE	SIZE
*1H01C13	519	*1C02C09	526
*2E03C02	519	*1H02C06	527
*2G01C13	519	*2B02C02	528
*3C01C06	519	*2E02C12	528
*2E03C20	519	*2C01C04	528
*3H01C22	520	*1C02C08	529
*2C02C02	520	*2B01C07	530
*3G01C06	520	*2C03C05	530
*2D01C14	520	*3F01C14	530
*2C02C07	521	*3H01C28	530
*1E01C02	521	*3F01C08	531
*2E03C19	521	*3C01C09	532
*1F01C12	522	*2H01C01	532
*3C02C02	522	*2C04C01	532
*3G02C07	522	*3G01C10	532
*2D04C10	522	*1H01C01	532
*2G02C06	522	*1G02C08	533
*1H01C02	522	*2D01C01	534
*3G02C13	522	*3H01C08	535
*1C01C01	523	*1D01C17	535
*2G01C10	523	*2E01C01	535
*2E03C18	523	*2E01C06	536
*3G02C01	523	*3F01C11	537
*3C01C03	523	*2C01C20	537
*2B02C11	523	*1H01C08	537
*1G02C03	524	*3H02C09	537
*2D02C20	524	*2E03C12	538
*2G02C09	524	*3G02C03	538
*2D05C03	524	*3C01C11	538
*2D04C08	524	*2G02C11	538
*2H02C03	525	*2D05C02	539
*2D01C03	525	*2G01C08	540
*1G02C11	525	*3H02C11	541
*2E03C05	525	*1F01C11	541
*3H01C05	525	*3F01C12	541
*2D02C02	525	*3B01C01	543
*2E03C06	525	*3G01C13	543
*1G02C10	526	*2G01C04	543
*1C01C07	526	*1H01C03	543
*2D02C10	526	*2D03C01	544

SAMPLES RANKED IN ASCENDING ORDER

SAMPLE	SIZE
*2H01C02	544
*2B02C03	545
*1F01C09	545
*2H02C05	546
*3F01C13	546
*3H01C13	546
*2C01C15	547
*3G01C01	548
*2E03C16	548
*1F01C07	549
*2D02C04	549
*3F01C07	549
*3C01C04	551
*2B01C09	551
*2E03C07	551
*1F01C08	552
*3B01C02	552
*3H01C09	552
*1D01C10	554
*3G02C14	555
*2D05C09	556
*1E01C05	556
*3C01C12	556
*2H01C08	557
*1D01C06	557
*2H01C05	557
*2E03C04	558
*2H01C04	559
*1D01C14	560
*2F01C05	561
*2E03C08	561
*1H01C06	562
*2C03C03	562
*2D04C01	564
*3C01C05	564
*3H01C06	565
*1F01C06	566
*2E03C21	567
*3H01C11	567
*3H01C02	568
*3C01C02	570
*2B02C01	571
*1D01C22	571
*2G01C11	572
*2C03C08	572
*2D04C04	572
*2E02C09	573
*3B01C05	573
*1D01C11	573
*1C02C04	575
*1D01C04	576
*1D01C09	577
*1E01C04	579
*2F01C07	581
*1D01C08	584
*1C02C06	588
*2D04C05	588
*3C01C07	594
*2B01C10	608
*1D01C21	611
*3C01C08	612
*1E01C03	612
*2D04C02	615
*2B01C05	618
*1D01C07	618
*3C01C10	620
*1D01C18	635
*2H01C03	638
*2E01C05	657

APPENDIX C

COMPUTER FILES AND PROGRAMS USED IN THE STUDY

FILE#	FILE NAME	SIZE	DESCRIPTION OF FILE
1	CORPUS	(394)	COMPLETE ENGLISH LANGUAGE SAMPLE.
2	GRADE8	(87)	SAMPLE TAKEN JUST FROM GRADE EIGHT.
3	GRADE9	(206)	" " " " GRADE NINE.
4	GRADE10	(102)	" " " " GRADE TEN.
5	AGRICULTURE	(18)	" " " " AGRICULTURE
6	COMMERCE	(35)	" " " " COMMERCE
7	ENGLISH	(70)	" " " " ENGLISH
8	HOMEC	(85)	" " " " HOME EC.
9	INDED	(54)	" " " " INDUSTRIAL ED.
10	MATH	(30)	" " " " MATHEMATICS
11	SCIENCE	(67)	" " " " SCIENCE
12	SOCIALS	(72)	" " " " SOCIALS
13	GRADE8.ENGLISH	(16)	GRADE-SUBJECT SAMPLE
14	GRADE8.HOMEC	(20)	" "
15	GRADE8.INDED	(9)	" "
16	GRADE8.MATH	(13)	" "
17	GRADE8.SCIENCE	(17)	" "
18	GRADE8.SOCIALS	(21)	" "
19	GRADE9.COMMERCE	(22)	" "
20	GRADE9.ENGLISH	(41)	" "
21	GRADE9.HOMEC	(66)	" "
22	GRADE9.INDED	(47)	" "
23	GRADE9.MATH	(7)	" "
24	GRADE9.SCIENCE	(23)	" "
25	GRADE9.SOCIALS	(14)	" "
26	GRADE10.COMMERCE	(14)	" "
27	GRADE10.ENGLISH	(15)	" "
28	GRADE10.MATH	(12)	" "
29	GRADE10.SCIENCE	(28)	" "
30	GRADE10.SOCIALS	(41)	" "
31	8E01	(7)	GRADE-SUBJECT-TEXT# SAMPLE
32	8E02	(9)	" " "
33	8H01	(20)	" " "
34	8I01	(9)	" " "
35	8M01	(13)	" " "
36	8SC01	(8)	" " "
37	8SC02	(11)	" " "
38	8SO01	(14)	" " "
39	8SO02	(7)	" " "
40	9C01	(11)	" " "

THE SIZE COLUMN REFERS TO MACHINE PAGES INSIDE THE COMPUTER, WHICH ARE APPROXIMATELY THE SAME AS TWO 8 X 11 PRINTED PAGES EACH.

FILE#	FILE NAME	SIZE	DESCRIPTION OF FILE
41	9C02	(13)	GRADE-SUBJECT-TEXT# SAMPLE
42	9E01	(19)	" " "
43	9E02	(7)	" " "
44	9E03	(9)	" " "
45	9E04	(9)	" " "
46	9H01	(18)	" " "
47	9H02	(19)	" " "
48	9H03	(13)	" " "
49	9H04	(10)	" " "
50	9H05	(9)	" " "
51	9I01	(12)	" " "
52	9I02	(14)	" " "
53	9I03	(23)	" " "
54	9M01	(7)	" " "
55	9SC01	(13)	" " "
56	9SC02	(11)	" " "
57	9SO01	(8)	" " "
58	9SO02	(6)	" " "
59	10C01	(6)	" " "
60	10C02	(8)	" " "
61	10E01	(12)	" " "
62	10E02	(5)	" " "
63	10M01	(12)	" " "
64	10SC01	(16)	" " "
65	10SC02	(14)	" " "
66	10SO01	(29)	" " "
67	10SO02	(12)	" " "
68	WRDSTAT.S	(4)	SENTENCE STATISTICS
69	WRDSTAT.O	(4)	" "
70	SPLIT1.S	(3)	PROGRAM TO BREAK 'CORPUS' INTO GDES.
71	SPLIT1.O	(2)	" " " " " "
72	ST.DEV.S	(1)	STANDARD DEVIATION PROGRAM
73	ST.DEV.O	(1)	" " "
74	TABL.B1.S	(1)	RANK TABLE (DESC. ORDER) PROGRAM
75	TABL.B1.O	(1)	" " " " "
76	TABL.B4.S	(1)	UNRANKED ASCENDING TABLE PROGRAM
77	TABL.B4.O	(1)	" " " "
78	SPLIT2.S	(3)	BREAKS GRADES INTO GRADE-SUBJECTS
79	SPLIT2.O	(3)	" " " " "
80	SPLIT3.S	(3)	BREAKS GRADE-SUBJECTS INTO TEXTS

THE SIZE COLUMN REFERS TO MACHINE PAGES INSIDE THE COMPUTER, WHICH ARE APPROXIMATELY THE SAME AS TWO 8 X 11 PRINTED PAGES EACH.

FILE#	FILE NAME	SIZE	DESCRIPTION OF FILE
81	SPLIT3.O	(3)	BREAKS GRADE-SUBJECTS INTO TEXTS
82	COUNTW.S	(7)	WORD COUNT PROGRAM
83	COUNTW.O	(7)	" " "
84	UNSORT.CORP.FREQS	(50)	CORPUS (TABL.B1, & TABL.B4 DATA)
85	UNSORT.GRD8.FREQS	(21)	GRADE8 "
86	UNSORT.GRD9.FREQS	(35)	GRADE9 "
87	UNSORT.GD10.FREQS	(24)	GRADE10 "
88	UNSORT.COMM.FREQS	(9)	COMMERCE "
89	UNSORT.ENGL.FREQS	(21)	ENGLISH "
90	UNSORT.HOME.FREQS	(17)	HOME EC. "
91	UNSORT.INDE.FREQS	(12)	INDUSTRIAL ED. "
92	UNSORT.MATH.FREQS	(6)	MATHEMATICS "
93	UNSORT.SCIE.FREQS	(15)	SCIENCE "
94	UNSORT.SOCI.FREQS	(20)	SOCIALS "
95	P1	(1)	P1 TO P37 FOLLOW 8E01 THRU' 10S002
96	P2	(1)	AND ARE THE INPUT DATA FOR PLOTT.S
97	P3	(1)	.
98	P4	(1)	.
99	P5	(1)	.
100	P6	(1)	.
101	P7	(1)	.
102	P8	(1)	.
103	P9	(1)	.
104	P10	(1)	.
105	P11	(1)	.
106	P12	(1)	.
107	P13	(1)	.
108	P14	(1)	.
109	P15	(1)	.
110	P16	(1)	.
111	P17	(1)	.
112	P18	(1)	.
113	P19	(1)	.
114	P20	(1)	.
115	P21	(1)	.
116	P22	(1)	.
117	P23	(1)	.
118	P24	(1)	.
119	P25	(1)	.
120	P26	(1)	.

THE SIZE COLUMN REFERS TO MACHINE PAGES INSIDE THE COMPUTER, WHICH ARE APPROXIMATELY THE SAME AS TWO 8 X 11 PRINTED PAGES EACH.

FILE#	FILE NAME	SIZE	DESCRIPTION OF FILE
121	P27	(1)	.
122	P28	(1)	.
123	P29	(1)	.
124	P30	(1)	.
125	P31	(1)	.
126	P32	(1)	.
127	P33	(1)	.
128	P34	(1)	.
129	P35	(1)	.
130	P36	(1)	.
131	P37	(1)	.
132	PG1	(1)	PG1 TO PG18 FOLLOW GRADE8.ENGLISH TO GRADE10.SOCIALS.
133	PG2	(1)	AND ARE THE INPUT DATA FOR PLOTGS.S
134	PG3	(1)	.
135	PG4	(1)	.
136	PG5	(1)	.
137	PG6	(1)	.
138	PG7	(1)	.
139	PG8	(1)	.
140	PG9	(1)	.
141	PG10	(1)	.
142	PG11	(1)	.
143	PG12	(1)	.
144	PG13	(1)	.
145	PG14	(1)	.
146	PG15	(1)	.
147	PG16	(1)	.
148	PG17	(1)	.
149	PG18	(1)	.
150	PS1	(1)	PS1 TO PS11 FOLLOW CORPUS TO
151	PS2	(1)	SOCIALS (EXCL. AGRICULTURE) AND ARE
152	PS3	(1)	INPUT DATA FOR PLOTG.S, & PLOTS.S
153	PS4	(1)	.
154	PS5	(1)	.
155	PS6	(1)	.
156	PS7	(1)	.
157	PS8	(1)	.
158	PS9	(1)	.
159	PS10	(1)	.
160	PS11	(1)	.

THE SIZE COLUMN REFERS TO MACHINE PAGES INSIDE THE COMPUTER, WHICH ARE APPROXIMATELY THE SAME AS TWO 8 X 11 PRINTED PAGES EACH.

FILE#	FILE NAME	SIZE	DESCRIPTION OF FILE
161	CORP.X2	(2)	CORPUS 'WORD' CHI-SQUARE TABLE
162	GRADES.X2	(2)	GRADES "
163	EIGHT.X2	(2)	GRADE8 "
164	NINE.X2	(2)	GRADE9 "
165	TEN.X2	(2)	GRADE10 "
166	CORSENT.X2	(1)	CORPUS 'SENTENCE' CHI-SQUARE TABLE
167	GDSSENT.X2	(1)	GRADES "
168	G8SENT.X2	(1)	GRADE8 "
169	G9SENT.X2	(1)	GRADE9 "
170	G10SENT.X2	(1)	GRADE10 "
171	LENGS	(1)	LENGTHS DATA FOR SENTENCE CHI-SQUARE
172	WORDS	(1)	WORDS DATA FOR WORDS CHI-SQUARE
173	CHIS.3	(1)	3 COLUMN CHI-SQUARE PROGRAM
174	CHIO.3	(1)	"
175	CHIS.4	(1)	3 COLUMN CHI-SQUARE PROGRAM (SENTS.)
176	CHIO.4	(1)	"
177	CHIS.7	(1)	7 COLUMN CHI-SQUARE PROGRAM
178	CHIO.7	(3)	"
179	CHIS.8	(1)	7 COLUMN CHI-SQUARE PROGRAM (SENTS.)
180	CHIO.8	(2)	"
181	COUNTW.O	(8)	WORD COUNT (OUTPUTS PLOT DATA TOO)
182	WRDSTAT.S	(4)	SENT. STATS. (OUTPUTS PLOT DATA TOO)
183	WRDSTAT.O	(4)	"
184	PLOTS.S	(1)	PLOT SENT. LENGTH DISTR. FOR SUBJS.
185	PLOTS.O	(1)	"
186	PLOTG.S	(1)	PLOT SENT. LENGTH DISTR. FOR CORPUS & GRADES
187	PLOTG.O	(1)	"
188	CORPUS.INDEX	(7)	TEXT INFORMATION FOR BOOKS
189	GRADES.INDEX	(7)	"
190	CORPUS.INTRODUCTIO	(3)	"
191	GRADES.INTRODUCTIO	(3)	"
192	GRADES.INTRO.INSER	(1)	"
193	PLOTT.S	(1)	PLOT SENT. LENGTH DISTR. FOR TEXTS
194	PLOTT.O	(1)	"
195	PLOTGS.S	(1)	PLOT SENT. LENGTH DISTR. FOR GRADE-SUBJECTS
196	PLOTGS.O	(1)	"
197	PW1	(3)	PW1 THRU' PW11 FOLLOW PS1 THRU' PS11
198	PW2	(2)	AND ARE INPUT DATA FOR PLOTGW.S
199	PW3	(3)	.
200	PW4	(1)	.

THE SIZE COLUMN REFERS TO MACHINE PAGES INSIDE THE COMPUTER, WHICH ARE APPROXIMATELY THE SAME AS TWO 8 X 11 PRINTED PAGES EACH.

FILE#	FILE NAME	SIZE	DESCRIPTION OF FILE
201	PW5	(1)	INPUT DATA FOR PLOTGW.S CONT'D
202	PW6	(1)	.
203	PW7	(2)	.
204	PW8	(1)	.
205	PW9	(1)	.
206	PW10	(1)	.
207	PW11	(1)	.
208	PLOTGW.S	(1)	PLOT WORD-FREQ-DISTR. (CORPUS, GRADES, & SUBJECTS)
209	TABL.B1W.S	(1)	VERSION OF TABL.B1.S TO GIVE DATA FOR PLOTGW.S
	TOTAL SIZE	(2,522)	

THE SIZE COLUMN REFERS TO MACHINE PAGES INSIDE THE COMPUTER, WHICH ARE APPROXIMATELY THE SAME AS TWO 8 X 11 PRINTED PAGES EACH.

APPENDIX D

ALPHABETICAL LISTING OF CORPUS VOCABULARY (SAMPLE)

FREQ	COUNT	WORD
501		
0.0043	1	ABBEY
0.0043	1	ABBOTS
0.0043	1	ABBREVIATED
0.0043	1	ABBREVIATING
0.0043	1	ABDICATE
0.0043	1	ABDICATED
0.0255	6	ABDOMEN
0.0085	2	ABE'S
0.0043	1	ABERDARES
0.0043	1	ABIDES
0.0213	5	ABILITIES
0.0936	22	ABILITY
0.3445	81	ABLE
0.0043	1	ABNER
0.0085	2	ABNORMAL
0.0043	1	ABNORMALITIES
0.0043	1	ABOARD
0.0043	1	ABODE
0.0043	1	ABOLISHED
0.0043	1	ABOUND
1.9693	463	ABOUT
0.4424	104	ABOVE
0.0043	1	ABRAHAM
0.0766	18	ABRASIVE
0.0468	11	ABRASIVES
0.0128	3	ABROAD
0.0043	1	ABRUPT
0.0043	1	ABRUPTLY
0.0043	1	ABSCURED
0.0213	5	ABSENCE
0.0128	3	ABSENT
0.0510	12	ABSOLUTE
0.0213	5	ABSOLUTELY
0.0468	11	ABSORB
0.0510	12	ABSORBED
0.0085	2	ABSORBENCY
0.0085	2	ABSORBENT
0.0128	3	ABSORBING
0.0255	6	ABSORBS
0.0085	2	ABSORPTION
0.0170	4	ABSTRACT
0.0043	1	ABSURDITY
0.0213	5	ABUNDANCE
0.0255	6	ABUNDANT
0.0085	2	ABUSES
0.0043	1	ABUTTED
0.0043	1	ACADEMY
0.0085	2	ACADIAN
0.0043	1	ACCELERATE
0.0043	1	ACCELERATED

FREQ	COUNT	WORD
551		
0.0085	2	ACCELERATING
0.0085	2	ACCELERATOR
0.0043	1	ACCELERATORS
0.0128	3	ACCENT
0.0043	1	ACCENTED
0.0638	15	ACCEPT
0.0213	5	ACCEPTABLE
0.0085	2	ACCEPTANCE
0.0383	9	ACCEPTED
0.0128	3	ACCEPTING
0.0043	1	ACCEPTS
0.0043	1	ACCESS
0.0043	1	ACCESSIBLE
0.0170	4	ACCESSORIES
0.0043	1	ACCESSORY
0.0298	7	ACCIDENT
0.0085	2	ACCIDENTALLY
0.0043	1	ACCIDENTALS
0.0383	9	ACCIDENTS
0.0043	1	ACCLIMATED
0.0043	1	ACCLIMATIZED
0.0043	1	ACCOMMODATE
0.0043	1	ACCOMMODATES
0.0043	1	ACCOMMODATIONS
0.0128	3	ACCOMPANIED
0.0170	4	ACCOMPANIES
0.0043	1	ACCOMPANIMENT
0.0043	1	ACCOMPANY
0.0213	5	ACCOMPANYING
0.0468	11	ACCOMPLISHED
0.0043	1	ACCOMPLISHES
0.0128	3	ACCOMPLISHMENT
0.0043	1	ACCOMPLISHMENTS
0.0085	2	ACCORD
0.0170	4	ACCORDANCE
0.1957	46	ACCORDING
0.0170	4	ACCORDINGLY
0.1999	47	ACCOUNT
0.0128	3	ACCOUNTANT
0.0128	3	ACCOUNTED
0.0043	1	ACCOUNTING
0.0681	16	ACCOUNTS
0.0128	3	ACCUMULATE
0.0043	1	ACCUMULATES
0.0043	1	ACCUMULATION
0.0808	19	ACCURACY
0.1531	36	ACCURATE
0.0766	18	ACCURATELY
0.0043	1	ACCUSED
0.0043	1	ACCUSING

APPENDIX E

RANK LISTING OF CORPUS VOCABULARY (SAMPLE)

CORPUS VOCABULARY RANK LIST

FREQ	COUNT	WORD
1		
7.4515	17519	THE
10.9295	8177	OF
13.6768	6459	AND
16.1952	5921	A
18.7136	5921	TO
20.8497	5022	IN
22.5140	3913	IS
23.4778	2266	THAT
24.4212	2218	IT
25.3421	2165	ARE
26.2527	2141	FOR
27.1417	2090	YOU
27.9375	1871	BE
28.6984	1789	AS
29.4002	1650	OR
30.0318	1485	WITH
30.6524	1459	ON
31.2015	1291	THIS
31.7315	1246	BY
32.2036	1110	WAS
32.6660	1087	HE
33.1185	1064	FROM
33.5681	1057	HAVE
34.0164	1054	AT
34.4384	992	WHICH
34.8352	933	ONE
35.2133	889	NOT
35.5778	857	CAN
35.9419	856	YOUR
36.3047	853	THEY
36.6590	833	WE
37.0104	826	HIS
37.3575	816	WILL
37.6888	779	IF
38.0172	772	AN
38.3149	700	WHEN
38.6045	681	ALL
38.8865	663	BUT
39.1583	639	THESE
39.4055	581	MAY
39.6475	569	THERE
39.8848	558	HAS
40.1196	552	I
40.3501	542	OTHER
40.5785	537	SOME
40.8035	529	MORE
41.0239	518	WERE
41.2408	510	HAD
41.4526	498	THEIR
41.6615	491	USED

FREQ	COUNT	WORD
51		
41.8690	488	MANY
42.0736	481	SO
42.2761	476	EACH
42.4738	465	TWO
42.6708	463	ABOUT
42.8575	439	SHOULD
43.0396	428	WHAT
43.2203	425	THAN
43.4007	424	BEEN
43.5806	423	INTO
43.7596	421	THEM
43.9370	417	USE
44.1118	411	MAKE
44.2845	406	DO
44.4559	403	UP
44.6261	400	SUCH
44.7962	400	THEN
44.9633	393	TIME
45.1275	386	ITS
45.2845	369	WOULD
45.4410	368	HOW
45.5967	366	NUMBER
45.7511	363	MADE
45.9034	358	OUT
46.0535	353	MOST
46.2028	351	ONLY
46.3512	349	NO
46.4967	342	MUST
46.6413	340	WATER
46.7740	312	ALSO
46.9050	308	FIRST
47.0352	306	VERY
47.1640	303	GOOD
47.2895	295	HIM
47.4112	286	SAME
47.5298	279	1
47.6404	260	COULD
47.7510	260	WHO
47.8595	255	ANY
47.9675	254	BECAUSE
48.0751	253	SEE
48.1793	245	LIKE
48.2810	239	MUCH
48.3814	236	PEOPLE
48.4809	234	CALLED
48.5804	234	2
48.6795	233	PLACE
48.7782	232	THROUGH
48.8769	232	WORK
48.9739	228	NEW

APPENDIX F

DESCENDING AND ASCENDING ORDER OF CORPUS VOCABULARY (SAMPLES)

THE CORPUS WITH RANK IN DESCENDING ORDER

RANK		X	FX	SUM FX	CUM% FX	FX*X	SUM FX*X	CUM% FX*X
	51	488	1	51	0.311	488	98437	41.869
	52	481	1	52	0.317	481	98918	42.073
	53	476	1	53	0.323	476	99394	42.276
	54	465	1	54	0.329	465	99859	42.473
	55	463	1	55	0.335	463	100322	42.670
	56	439	1	56	0.341	439	100761	42.857
	57	428	1	57	0.347	428	101189	43.039
	58	425	1	58	0.354	425	101614	43.220
	59	424	1	59	0.360	424	102038	43.400
	60	423	1	60	0.366	423	102461	43.580
	61	421	1	61	0.372	421	102882	43.759
	62	417	1	62	0.378	417	103299	43.937
	63	411	1	63	0.384	411	103710	44.111
	64	406	1	64	0.390	406	104116	44.284
	65	403	1	65	0.396	403	104519	44.455
66 -	67	400	2	67	0.408	800	105319	44.796
	68	393	1	68	0.415	393	105712	44.963
	69	386	1	69	0.421	386	106098	45.127
	70	369	1	70	0.427	369	106467	45.284
	71	368	1	71	0.433	368	106835	45.440
	72	366	1	72	0.439	366	107201	45.596
	73	363	1	73	0.445	363	107564	45.750
	74	358	1	74	0.451	358	107922	45.903
	75	353	1	75	0.457	353	108275	46.053
	76	351	1	76	0.463	351	108626	46.202
	77	349	1	77	0.469	349	108975	46.351
	78	342	1	78	0.475	342	109317	46.496
	79	340	1	79	0.482	340	109657	46.641
	80	312	1	80	0.488	312	109969	46.773
	81	308	1	81	0.494	308	110277	46.904
	82	306	1	82	0.500	306	110583	47.035
	83	303	1	83	0.506	303	110886	47.163
	84	295	1	84	0.512	295	111181	47.289
	85	286	1	85	0.518	286	111467	47.411
	86	279	1	86	0.524	279	111746	47.529
87 -	88	260	2	88	0.536	520	112266	47.750
	89	255	1	89	0.543	255	112521	47.859
	90	254	1	90	0.549	254	112775	47.967
	91	253	1	91	0.555	253	113028	48.074
	92	245	1	92	0.561	245	113273	48.179
	93	239	1	93	0.567	239	113512	48.280
	94	236	1	94	0.573	236	113748	48.381
95 -	96	234	2	96	0.585	468	114216	48.580
	97	233	1	97	0.591	233	114449	48.679
98 -	99	232	2	99	0.603	464	114913	48.876
	100	228	1	100	0.610	228	115141	48.973
	101	223	1	101	0.616	223	115364	49.068
	102	220	1	102	0.622	220	115584	49.162
	103	217	1	103	0.628	217	115801	49.254
104 -	105	216	2	105	0.640	432	116233	49.438

THE CORPUS IN ASCENDING ORDER

X	FX	SUM FX	CUM% FX	FX*X	SUM FX*X	CUM% FX*X
1	7098	7098	43.267	7098	7098	3.019
2	2418	9516	58.007	4836	11934	5.076
3	1240	10756	65.565	3720	15654	6.658
4	854	11610	70.771	3416	19070	8.111
5	632	12242	74.624	3160	22230	9.455
6	453	12695	77.385	2718	24948	10.611
7	387	13082	79.744	2709	27657	11.764
8	322	13404	81.707	2576	30233	12.859
9	251	13655	83.237	2259	32492	13.820
10	208	13863	84.505	2080	34572	14.705
11	189	14052	85.657	2079	36651	15.589
12	156	14208	86.608	1872	38523	16.385
13	123	14331	87.357	1599	40122	17.065
14	138	14469	88.199	1932	42054	17.887
15	113	14582	88.887	1695	43749	18.608
16	100	14682	89.497	1600	45349	19.289
17	98	14780	90.094	1666	47015	19.997
18	98	14878	90.692	1764	48779	20.748
19	58	14936	91.045	1102	49881	21.216
20	68	15004	91.460	1360	51241	21.795
21	51	15055	91.771	1071	52312	22.250
22	63	15118	92.155	1386	53698	22.840
23	36	15154	92.374	828	54526	23.192
24	44	15198	92.642	1056	55582	23.641
25	37	15235	92.868	925	56507	24.034
26	44	15279	93.136	1144	57651	24.521
27	39	15318	93.374	1053	58704	24.969
28	34	15352	93.581	952	59656	25.374
29	33	15385	93.782	957	60613	25.781
30	38	15423	94.014	1140	61753	26.266
31	27	15450	94.178	837	62590	26.622
32	30	15480	94.361	960	63550	27.030
33	21	15501	94.489	693	64243	27.325
34	39	15540	94.727	1326	65569	27.889
35	40	15580	94.971	1400	66969	28.484
36	31	15611	95.160	1116	68085	28.959
37	26	15637	95.318	962	69047	29.368
38	25	15662	95.471	950	69997	29.772
39	21	15683	95.599	819	70816	30.121
40	26	15709	95.757	1040	71856	30.563
41	18	15727	95.867	738	72594	30.877
42	21	15748	95.995	882	73476	31.252
43	24	15772	96.141	1032	74508	31.691
44	17	15789	96.245	748	75256	32.009
45	13	15802	96.324	585	75841	32.258
46	18	15820	96.434	828	76669	32.610
47	10	15830	96.495	470	77139	32.810
48	12	15842	96.568	576	77715	33.055
49	8	15850	96.617	392	78107	33.222

APPENDIX G

SENTENCE LENGTH DISTRIBUTION OF THE CORPUS (SAMPLE)

SENTENCE-LENGTH DISTRIBUTION OF
THE CORPUS

LENGTH	REPETITIONS	CUM. SENT	ACCUM WORDS	% WORDS
1	36	36	36	0.02
2	86	122	208	0.09
3	100	222	508	0.22
4	176	398	1212	0.52
5	283	681	2627	1.12
6	313	994	4505	1.92
7	415	1409	7410	3.15
8	471	1880	11178	4.75
9	522	2402	15876	6.75
10	581	2983	21686	9.22
11	577	3560	28033	11.92
12	616	4176	35425	15.07
13	623	4799	43524	18.51
14	632	5431	52372	22.28
15	655	6086	62197	26.46
16	634	6720	72341	30.77
17	584	7304	82269	34.99
18	580	7884	92709	39.43
19	505	8389	102304	43.51
20	471	8860	111724	47.52
21	432	9292	120796	51.38
22	424	9716	130124	55.35
23	384	10100	138956	59.10
24	343	10443	147188	62.61
25	294	10737	154538	65.73
26	266	11003	161454	68.67
27	252	11255	168258	71.57
28	239	11494	174950	74.41
29	192	11686	180518	76.78
30	193	11879	186308	79.25
31	171	12050	191609	81.50
32	124	12174	195577	83.19
33	109	12283	199174	84.72
34	103	12386	202676	86.21
35	81	12467	205511	87.41
36	75	12542	208211	88.56
37	59	12601	210394	89.49
38	43	12644	212028	90.19
39	50	12694	213978	91.02
40	48	12742	215898	91.83
41	50	12792	217948	92.70
42	36	12828	219460	93.35
43	24	12852	220492	93.79
44	22	12874	221460	94.20
45	19	12893	222315	94.56
46	23	12916	223373	95.01

APPENDIX H

GRAPHS OF SENTENCE LENGTH DISTRIBUTION

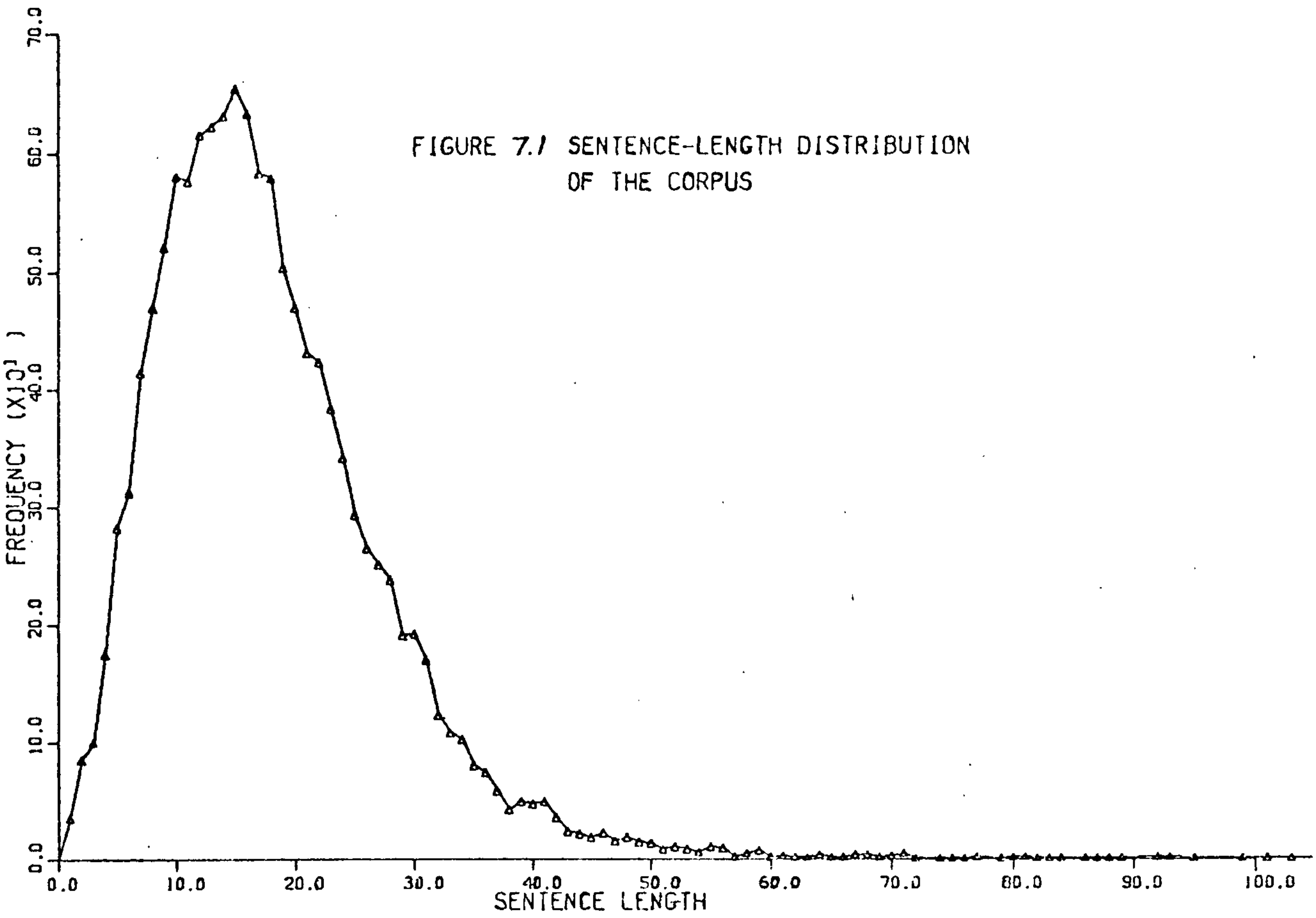


FIGURE 7.1 SENTENCE-LENGTH DISTRIBUTION OF THE CORPUS

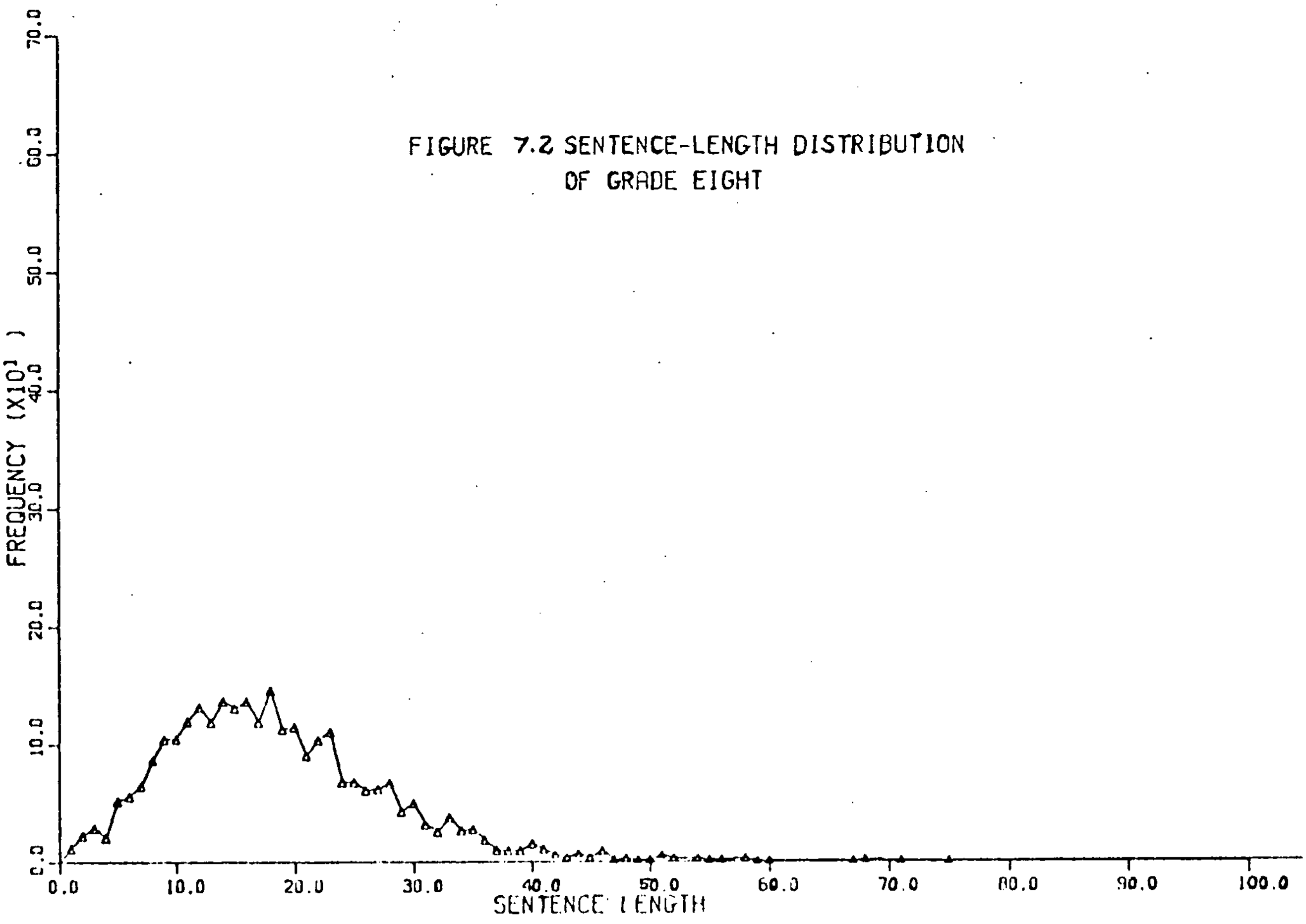


FIGURE 7.2 SENTENCE-LENGTH DISTRIBUTION OF GRADE EIGHT

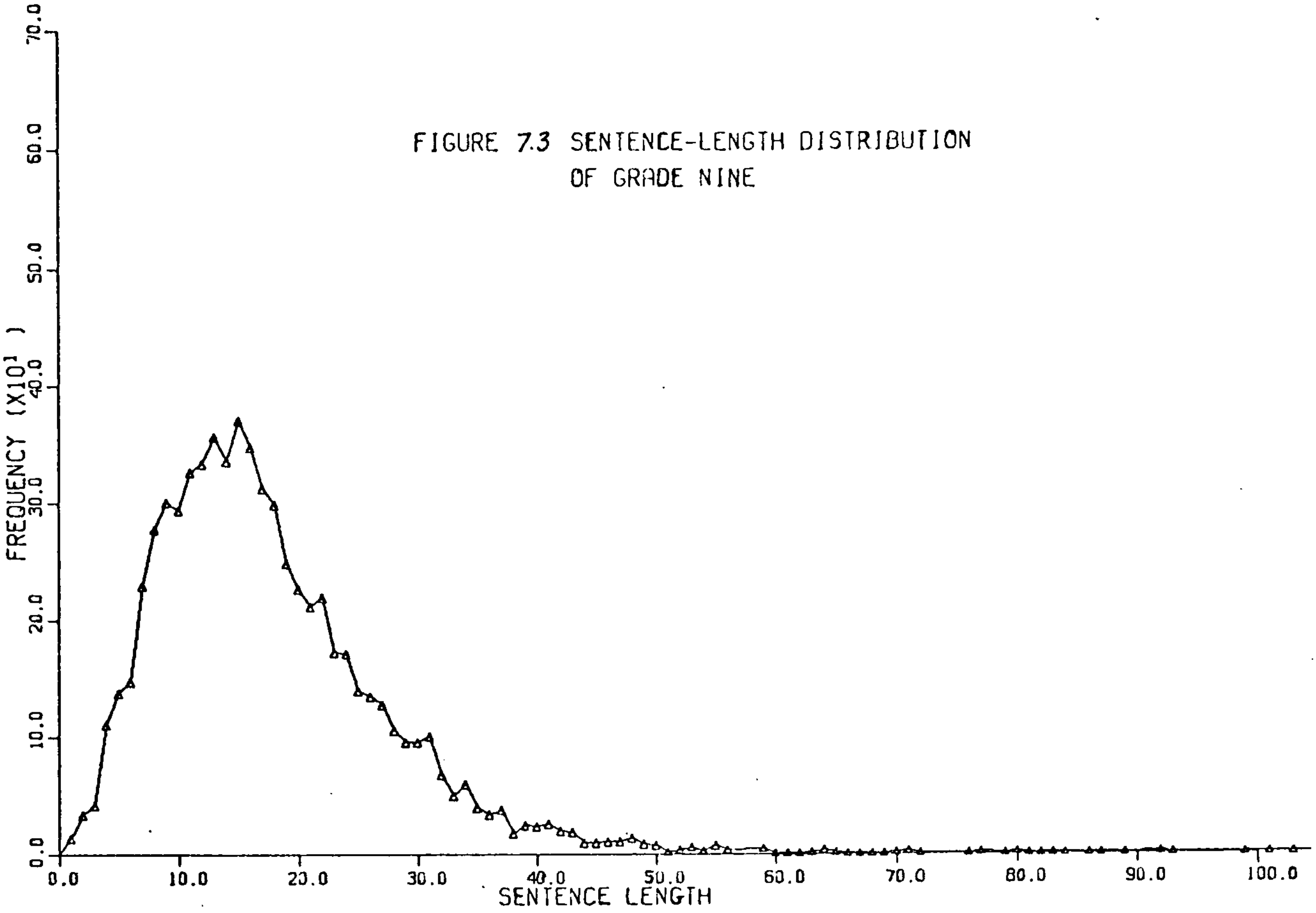


FIGURE 7.3 SENTENCE-LENGTH DISTRIBUTION OF GRADE NINE

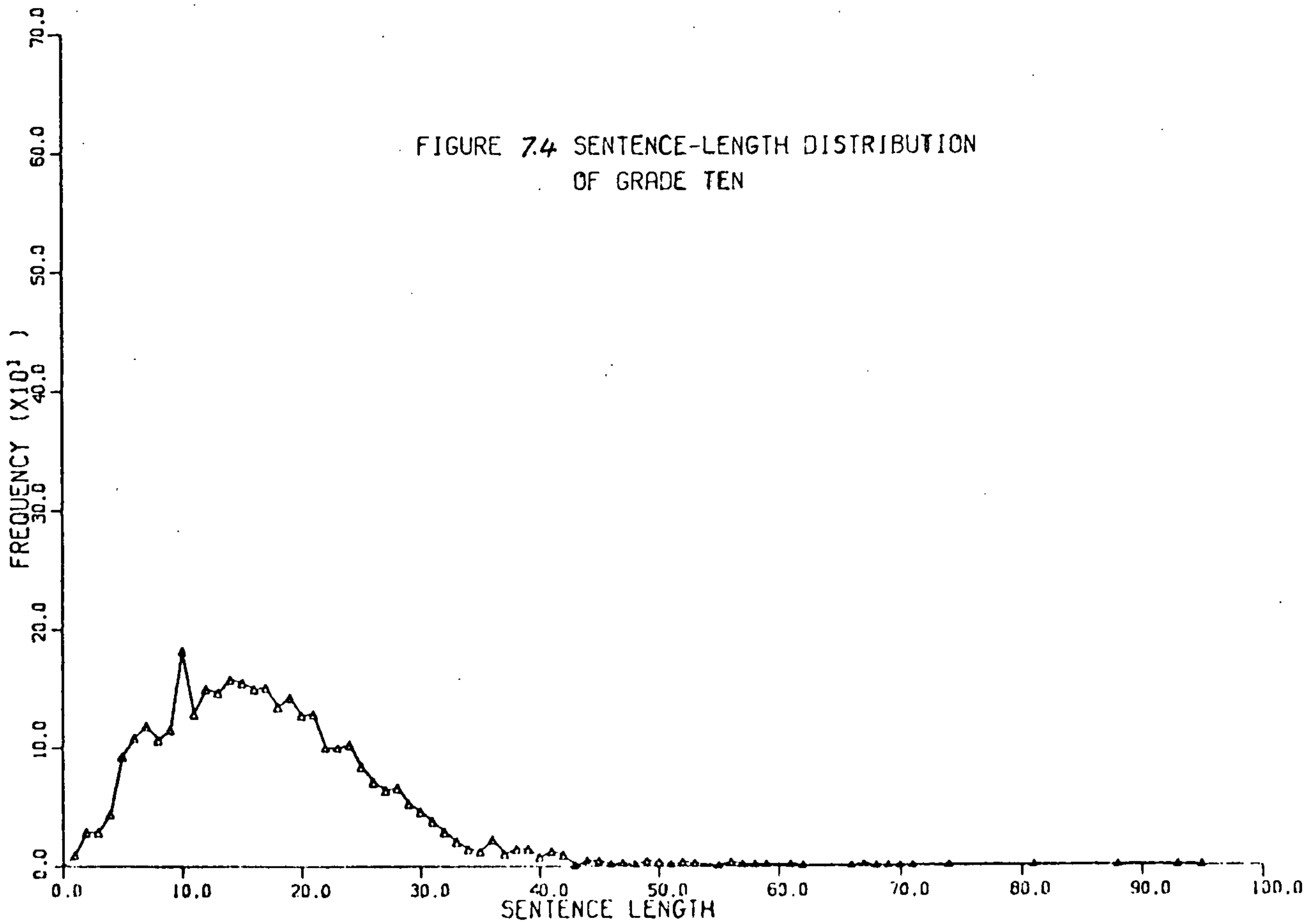


FIGURE 7.4 SENTENCE-LENGTH DISTRIBUTION OF GRADE TEN

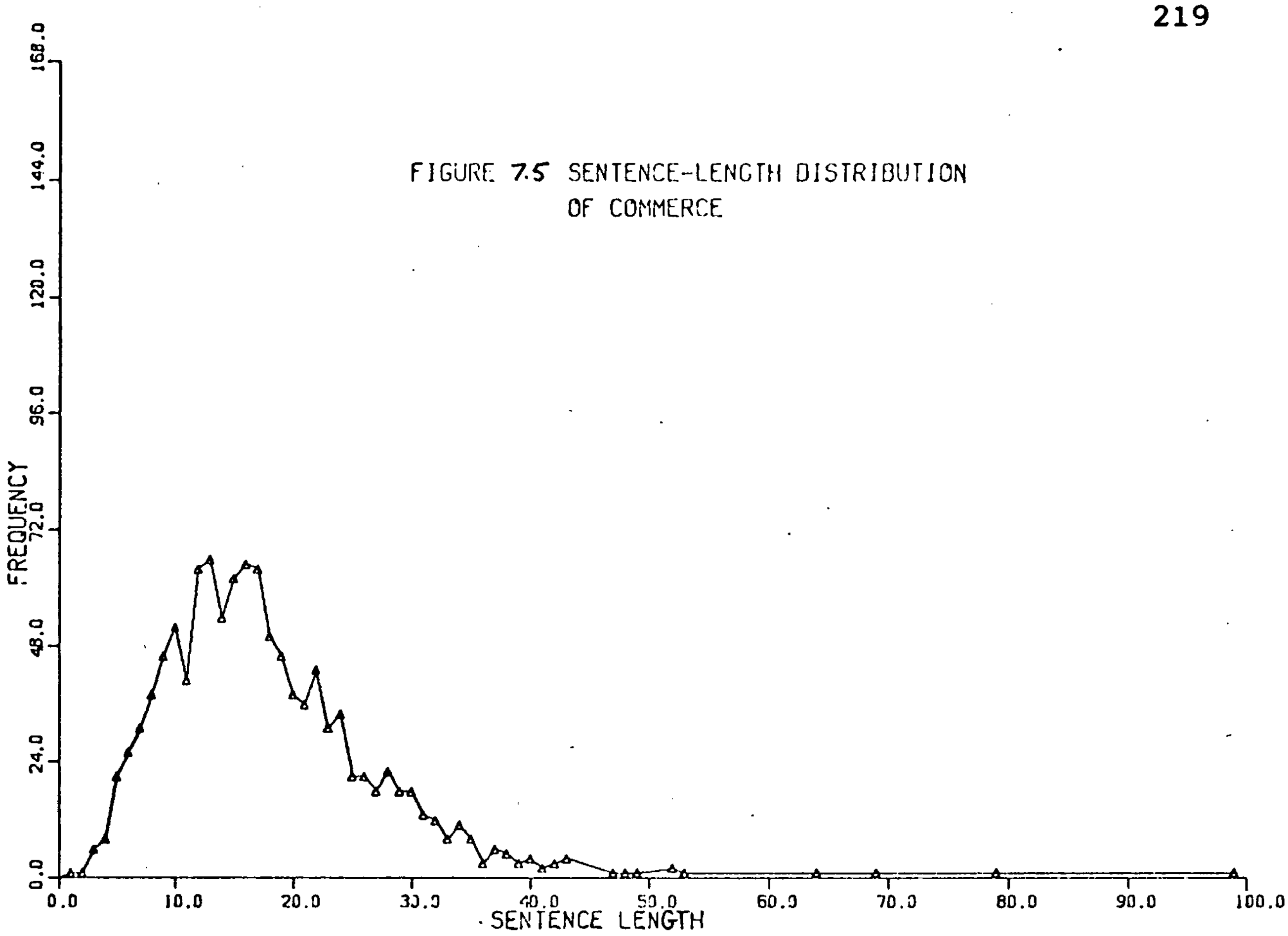


FIGURE 7.5 SENTENCE-LENGTH DISTRIBUTION
OF COMMERCE
FREQUENCY
SENTENCE LENGTH

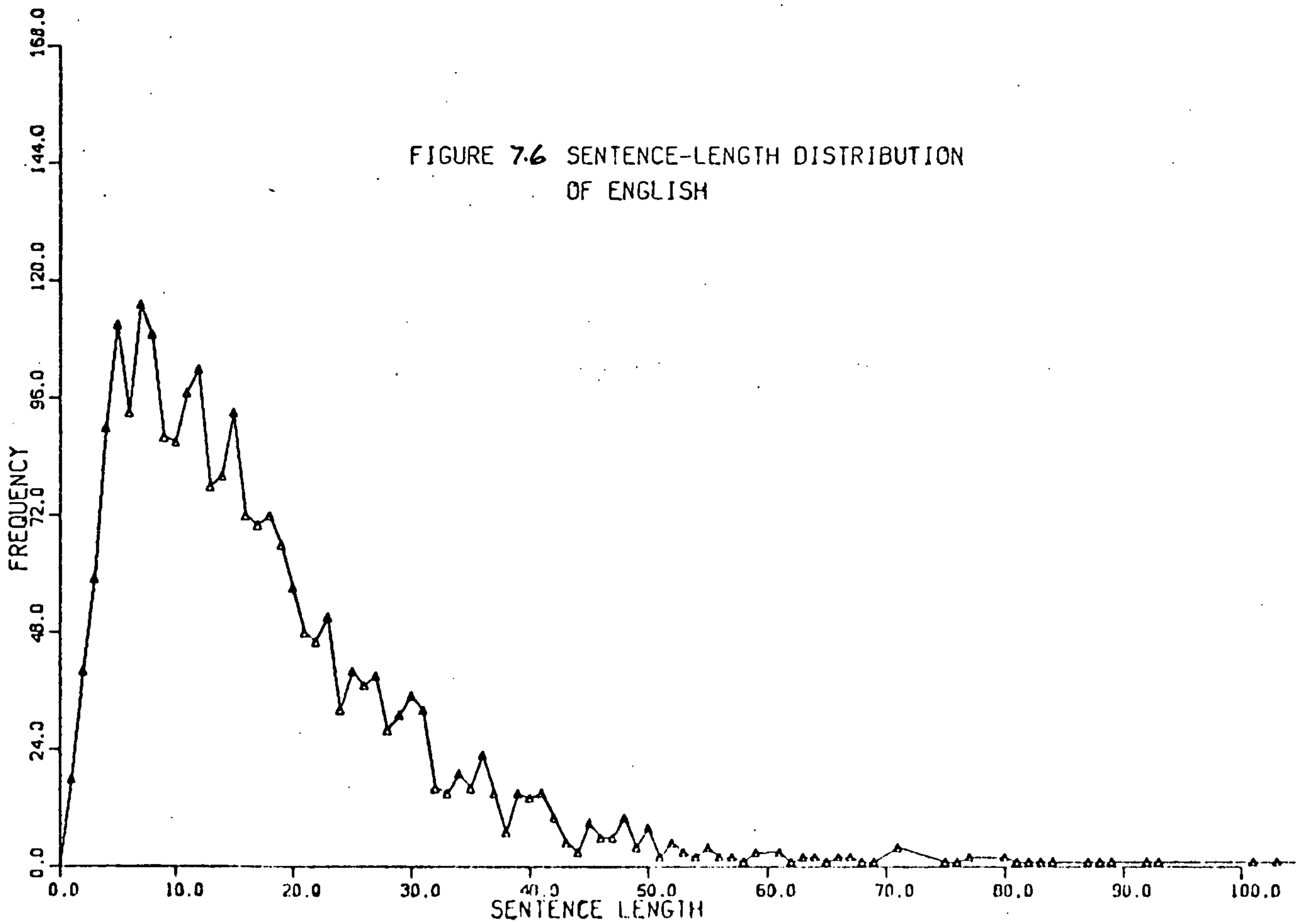


FIGURE 7.6 SENTENCE-LENGTH DISTRIBUTION
OF ENGLISH
FREQUENCY
SENTENCE LENGTH

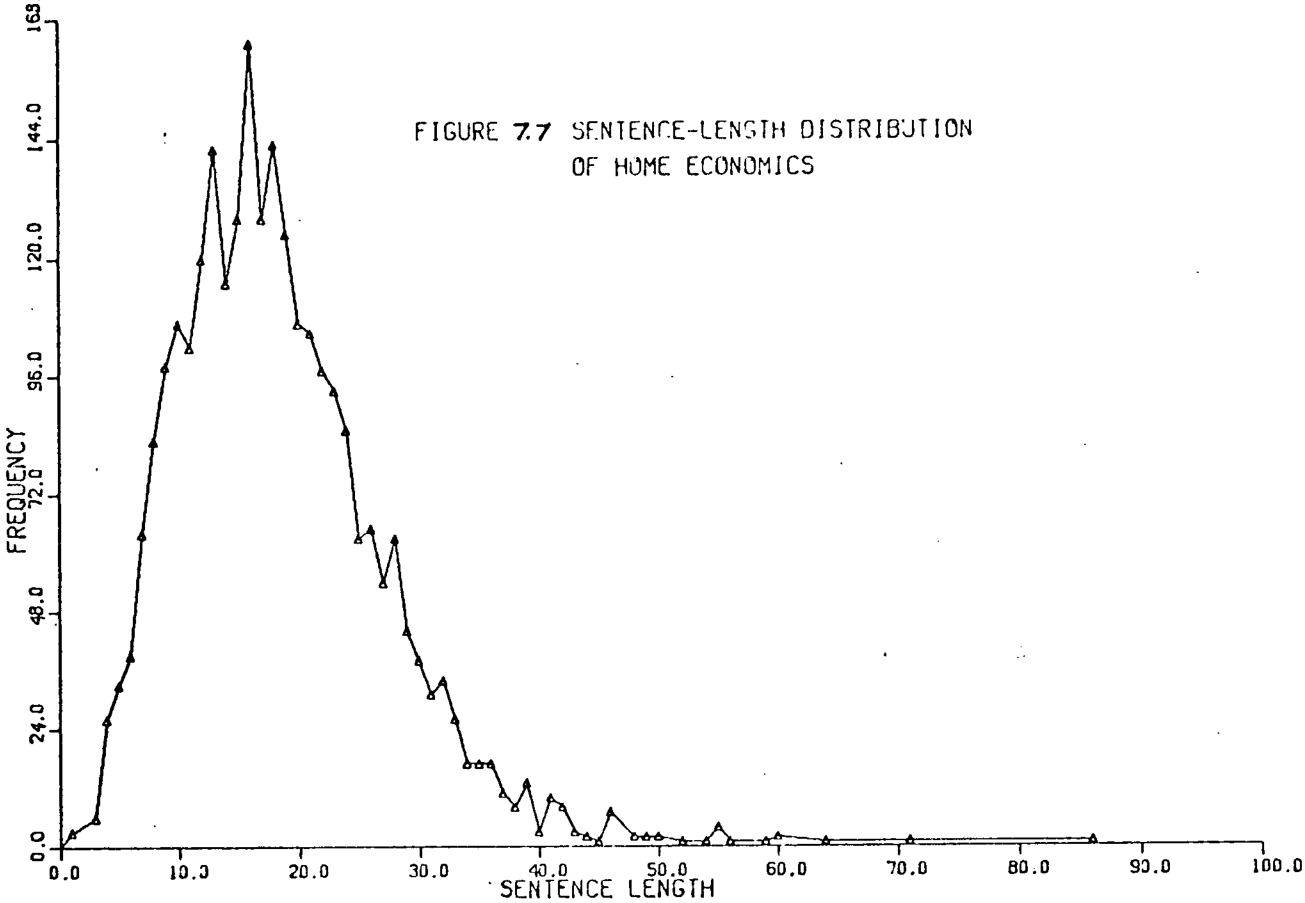


FIGURE 7.7 SENTENCE-LENGTH DISTRIBUTION
OF HOME ECONOMICS
FREQUENCY
SENTENCE LENGTH
0.0
24.0
48.0
72.0
96.0
120.0
144.0
168.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0

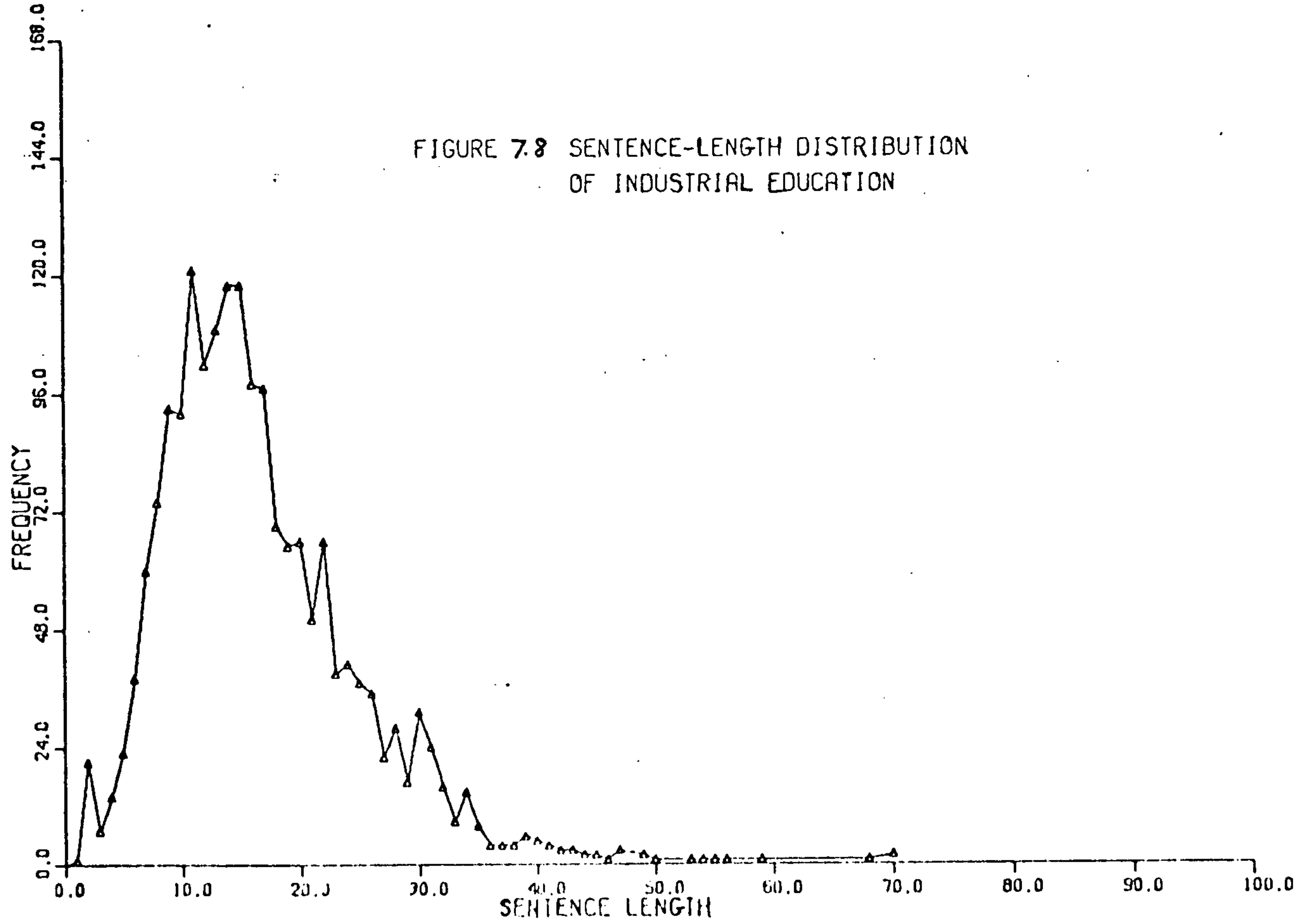


FIGURE 7.8 SENTENCE-LENGTH DISTRIBUTION
OF INDUSTRIAL EDUCATION
FREQUENCY
SENTENCE LENGTH
0.0
24.0
48.0
72.0
96.0
120.0
144.0
168.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0

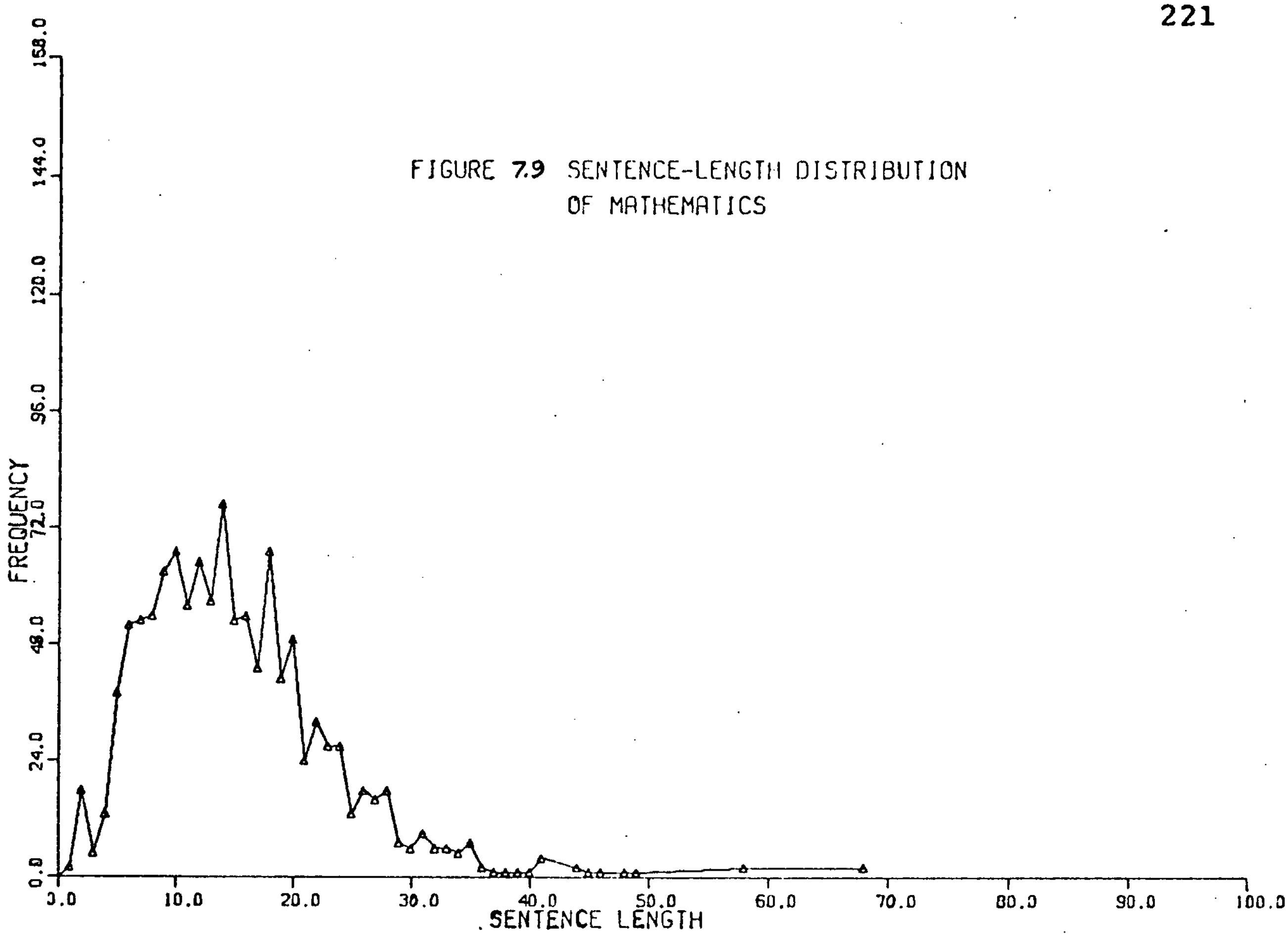


FIGURE 7.9 SENTENCE-LENGTH DISTRIBUTION OF MATHEMATICS

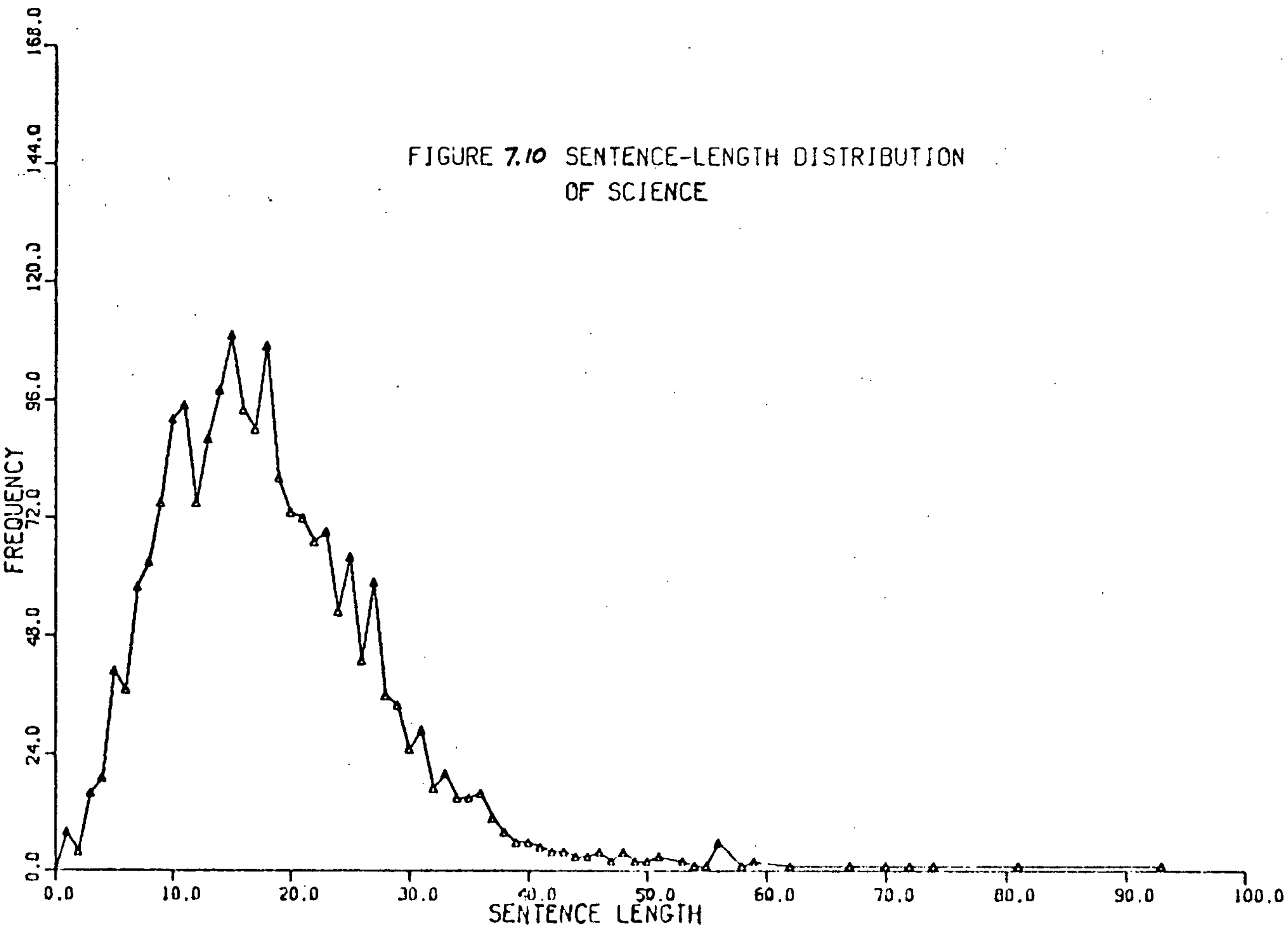


FIGURE 7.10 SENTENCE-LENGTH DISTRIBUTION OF SCIENCE

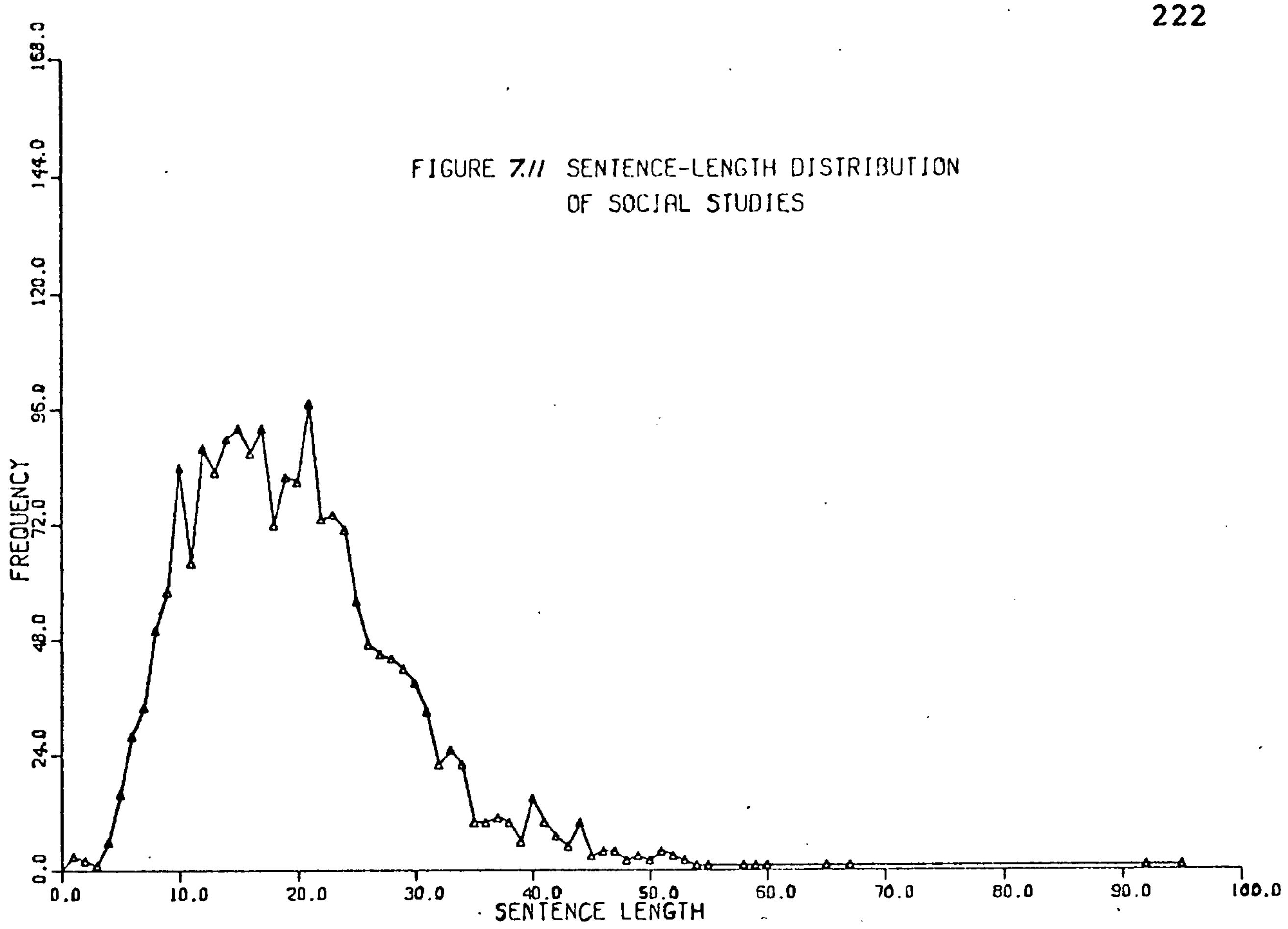


FIGURE 7.11 SENTENCE-LENGTH DISTRIBUTION
OF SOCIAL STUDIES
FREQUENCY
SENTENCE LENGTH
0.0
24.0
48.0
72.0
96.0
120.0
144.0
168.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0

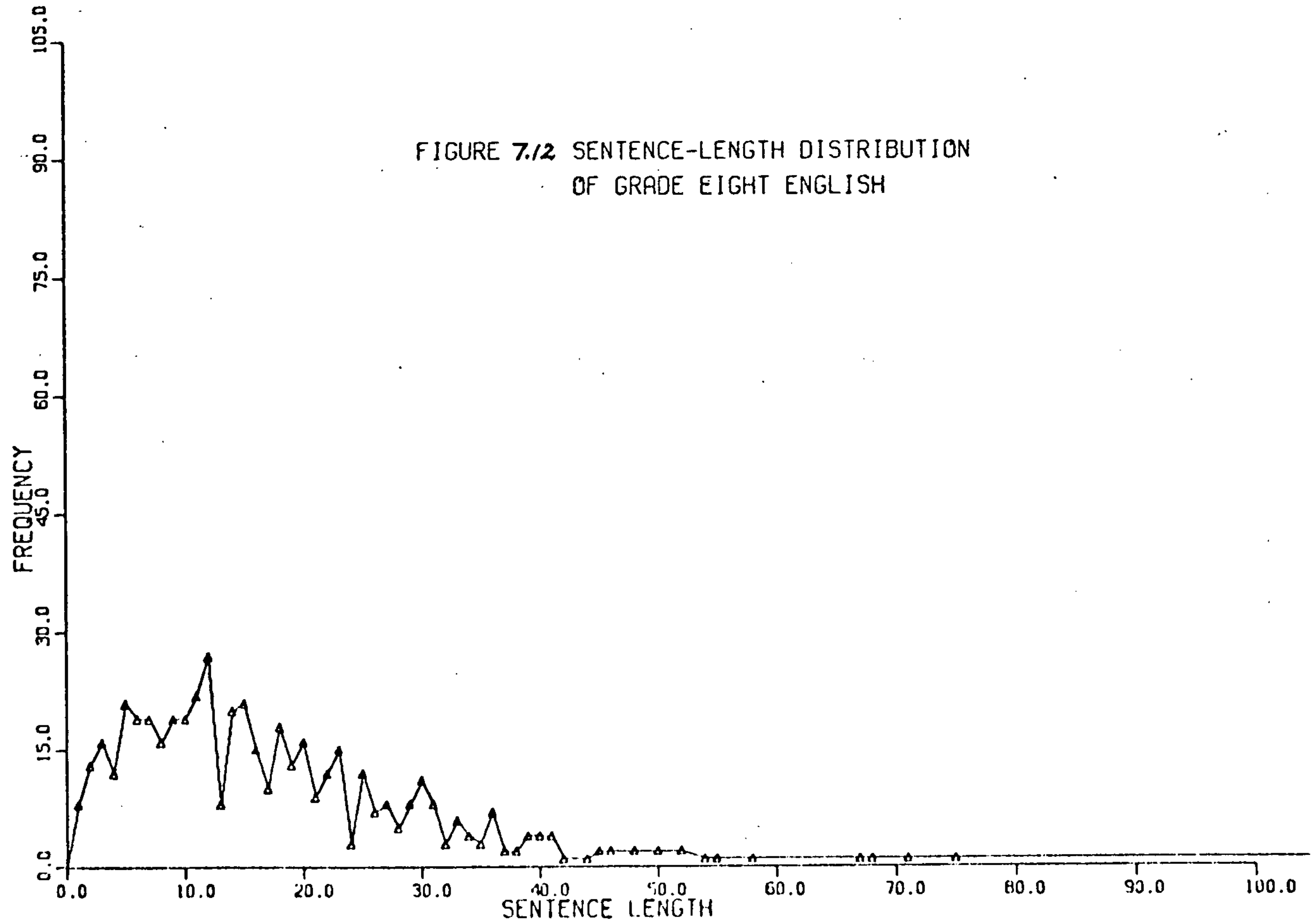


FIGURE 7.12 SENTENCE-LENGTH DISTRIBUTION
OF GRADE EIGHT ENGLISH
FREQUENCY
SENTENCE LENGTH
0.0
15.0
30.0
45.0
60.0
75.0
90.0
105.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0

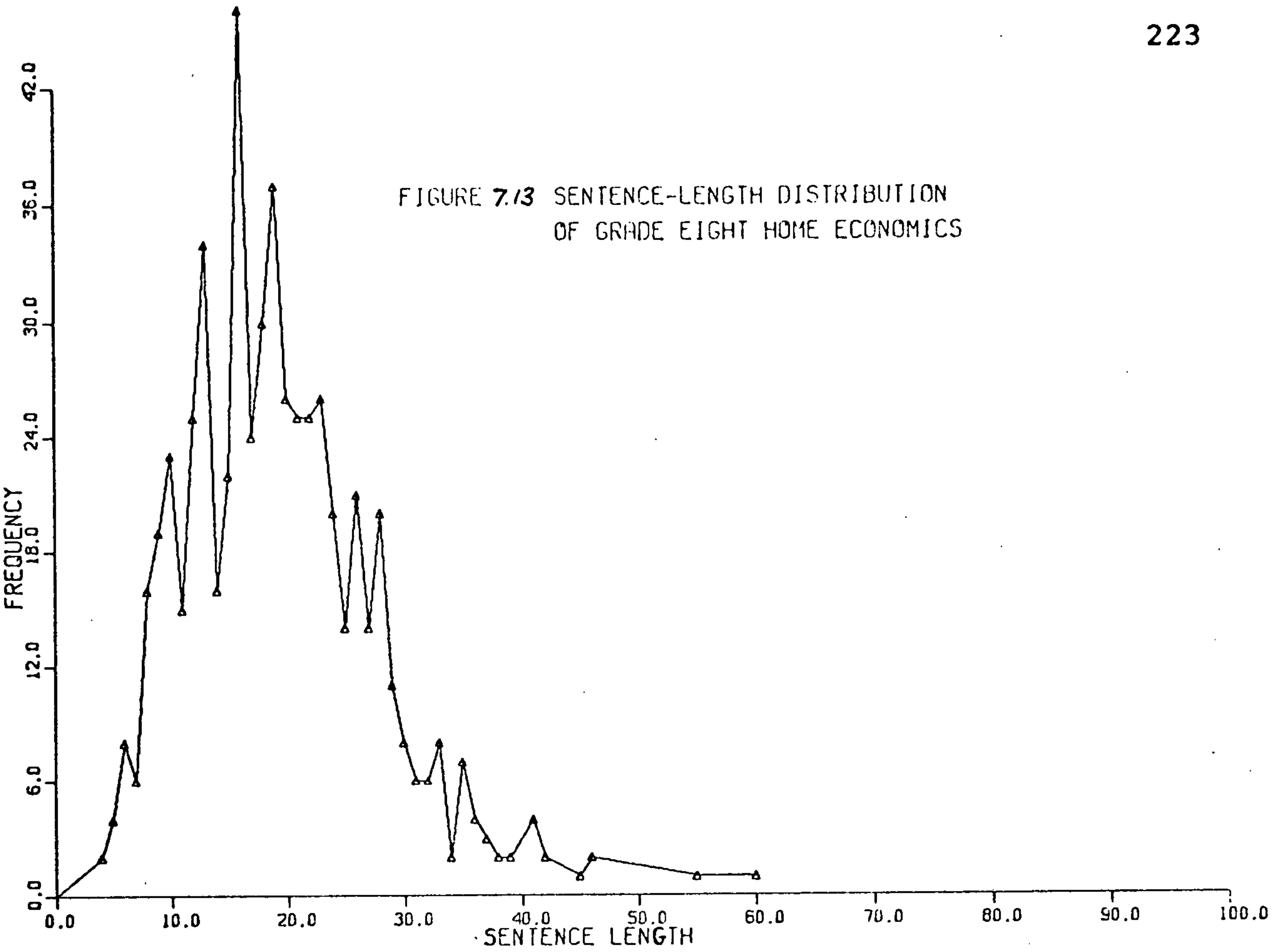


FIGURE 7.13 SENTENCE-LENGTH DISTRIBUTION OF GRADE EIGHT HOME ECONOMICS

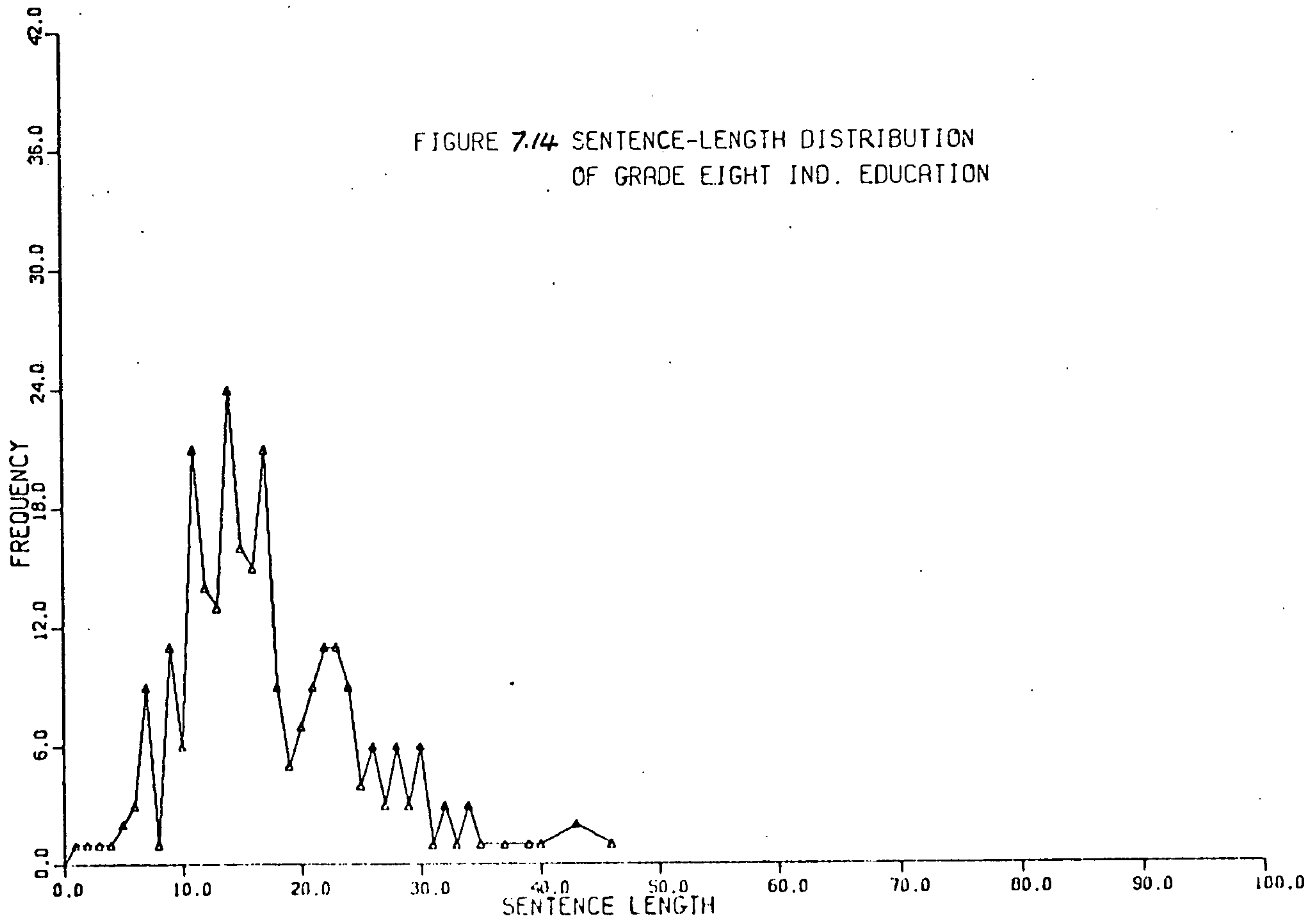


FIGURE 7.14 SENTENCE-LENGTH DISTRIBUTION OF GRADE EIGHT IND. EDUCATION

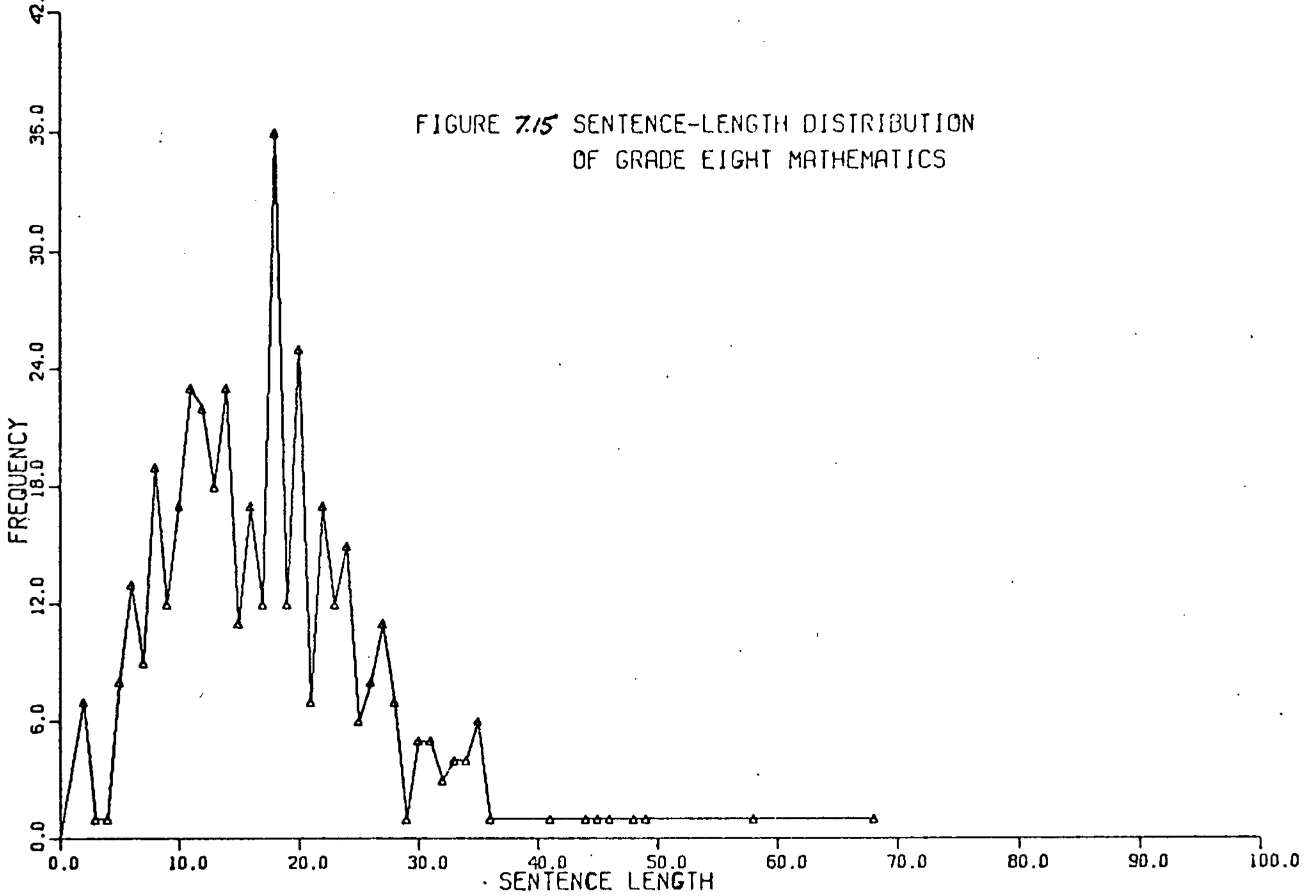


FIGURE 7.15 SENTENCE-LENGTH DISTRIBUTION OF GRADE EIGHT MATHEMATICS

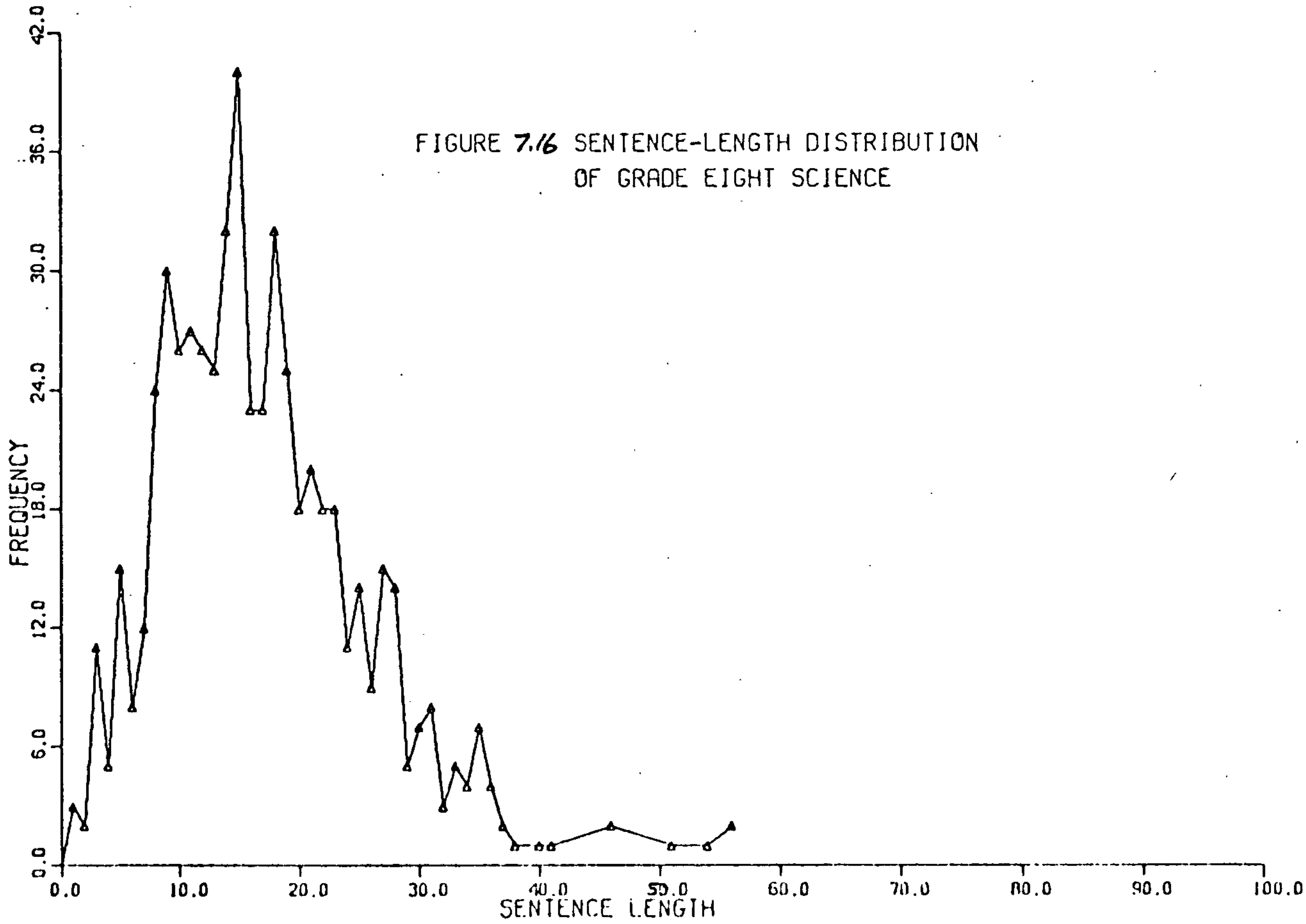


FIGURE 7.16 SENTENCE-LENGTH DISTRIBUTION OF GRADE EIGHT SCIENCE

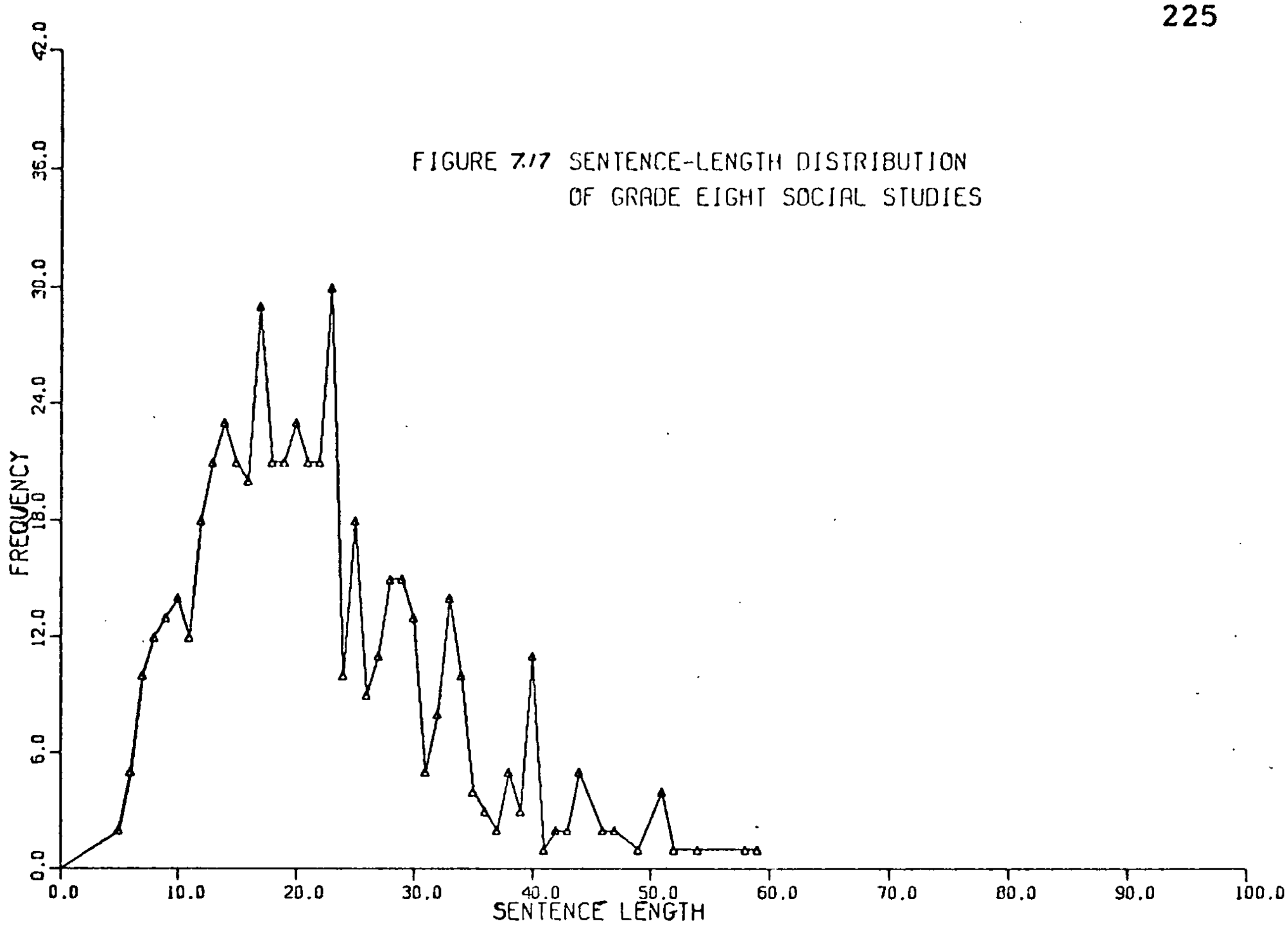


FIGURE 7.17 SENTENCE-LENGTH DISTRIBUTION OF GRADE EIGHT SOCIAL STUDIES
FREQUENCY
SENTENCE LENGTH
0.0
6.0
12.0
18.0
24.0
30.0
36.0
42.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0

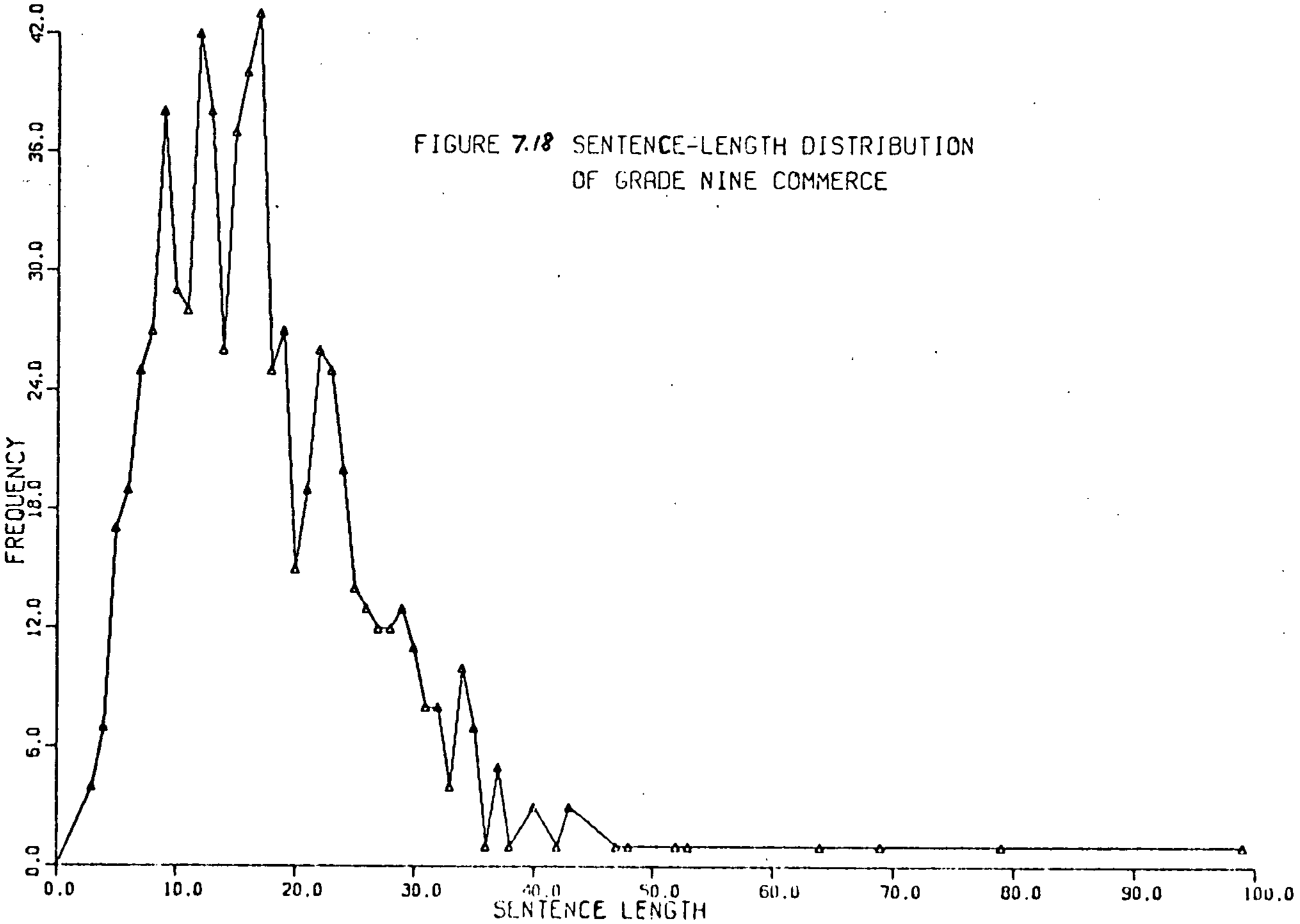


FIGURE 7.18 SENTENCE-LENGTH DISTRIBUTION OF GRADE NINE COMMERCE
FREQUENCY
SENTENCE LENGTH
0.0
6.0
12.0
18.0
24.0
30.0
36.0
42.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0

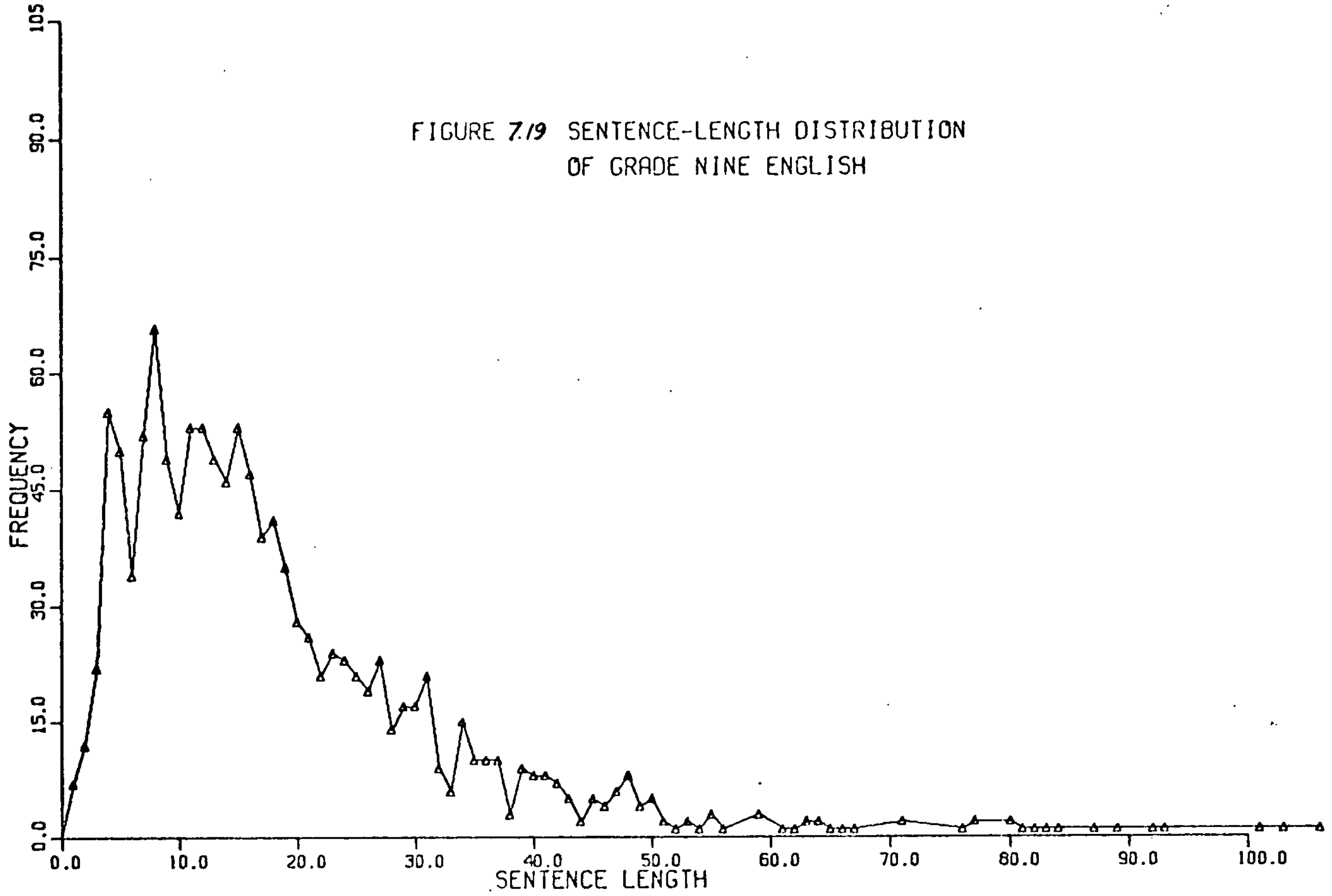


FIGURE 7.19 SENTENCE-LENGTH DISTRIBUTION OF GRADE NINE ENGLISH

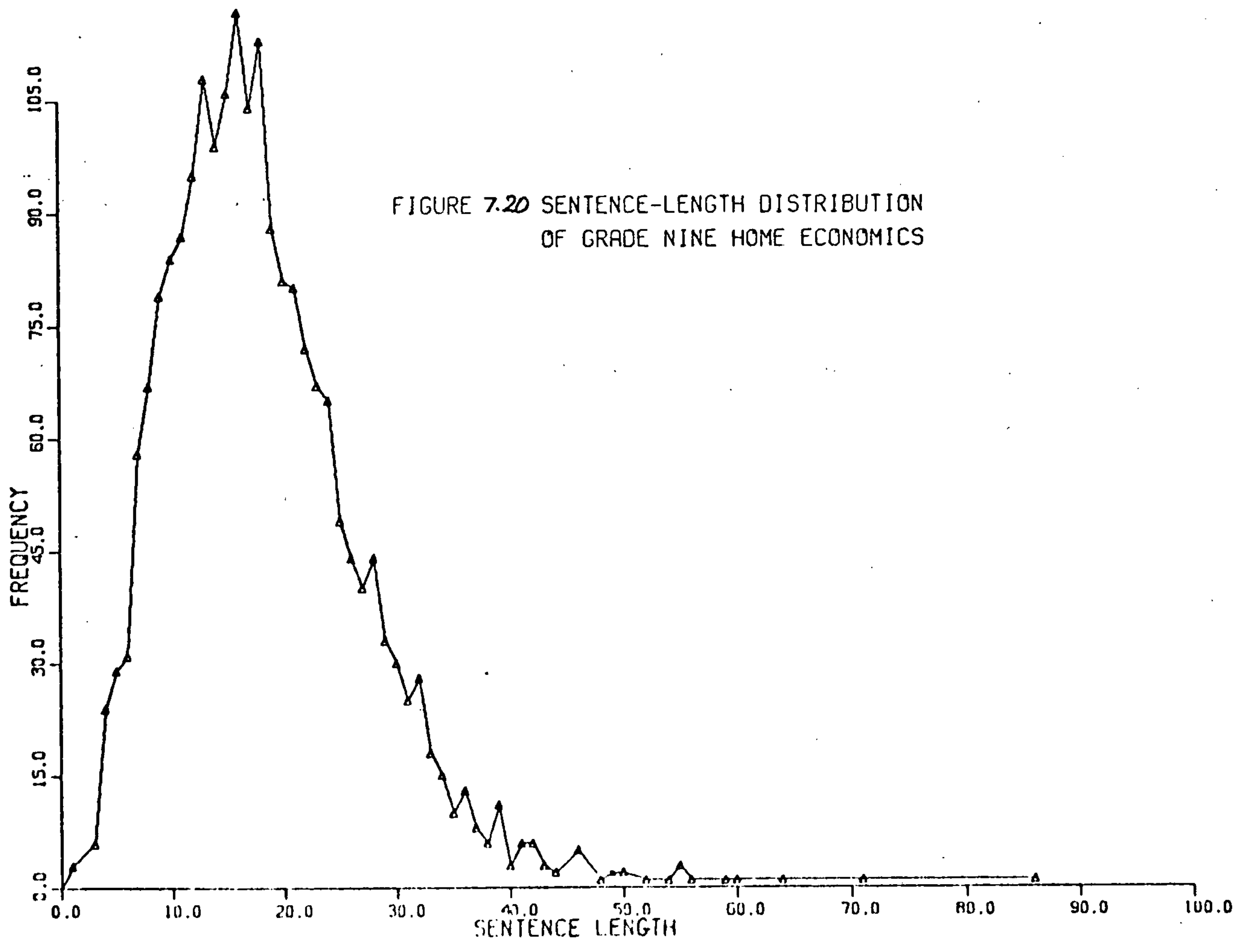


FIGURE 7.20 SENTENCE-LENGTH DISTRIBUTION OF GRADE NINE HOME ECONOMICS

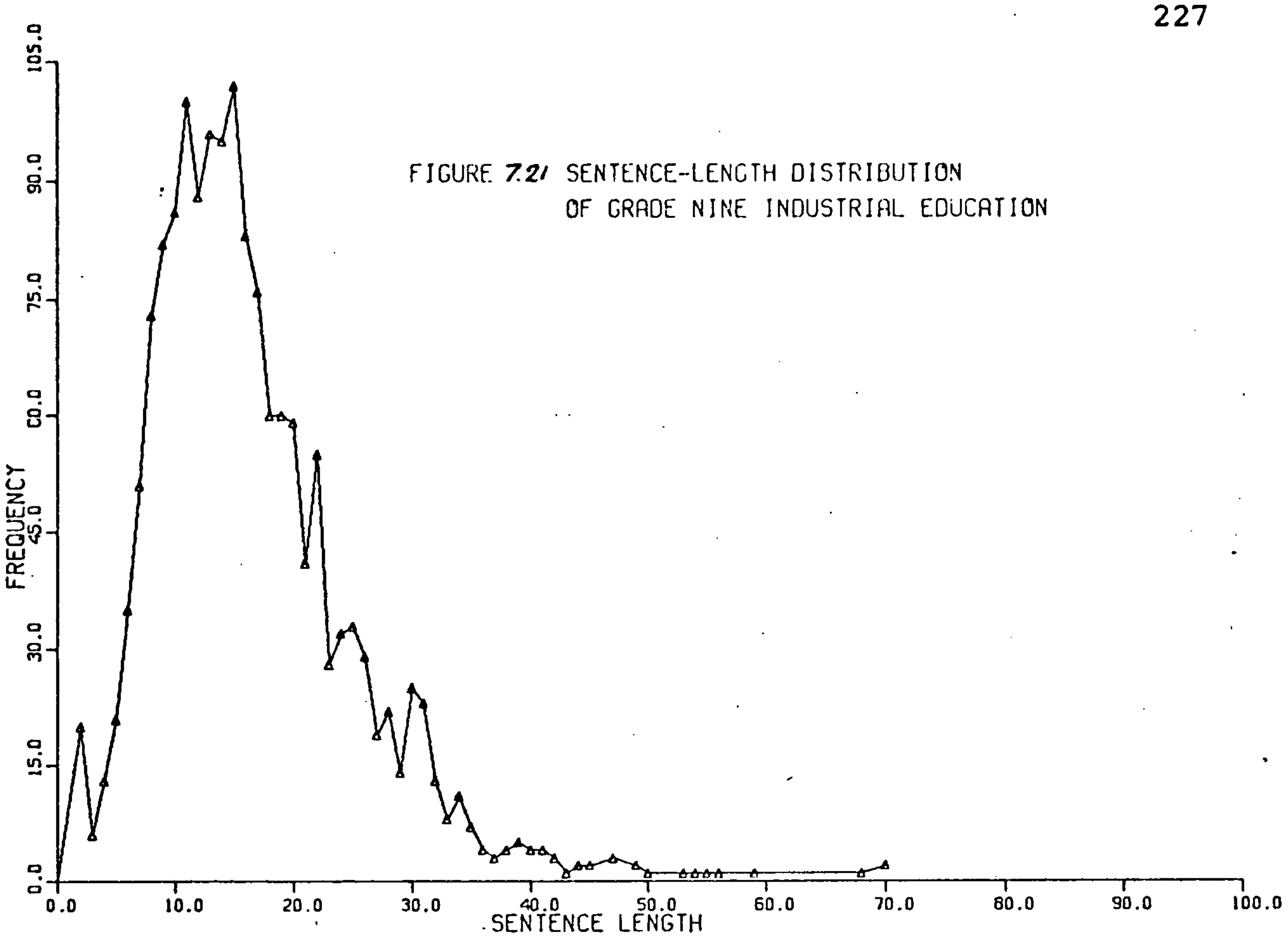


FIGURE 7.21 SENTENCE-LENGTH DISTRIBUTION
OF GRADE NINE INDUSTRIAL EDUCATION
FREQUENCY
SENTENCE LENGTH
0.0
15.0
30.0
45.0
60.0
75.0
90.0
105.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0

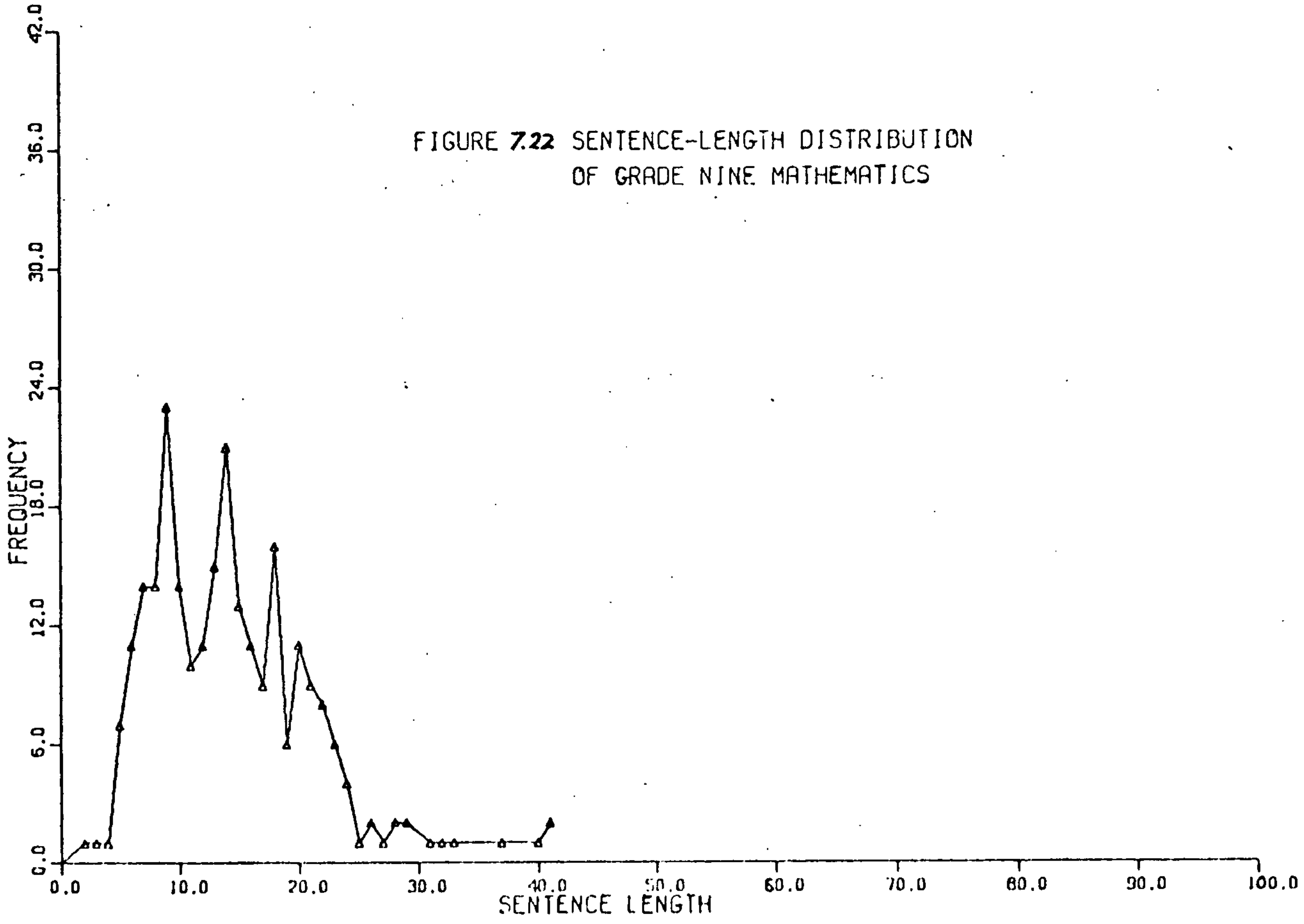


FIGURE 7.22 SENTENCE-LENGTH DISTRIBUTION
OF GRADE NINE MATHEMATICS
FREQUENCY
SENTENCE LENGTH
0.0
6.0
12.0
18.0
24.0
30.0
36.0
42.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0

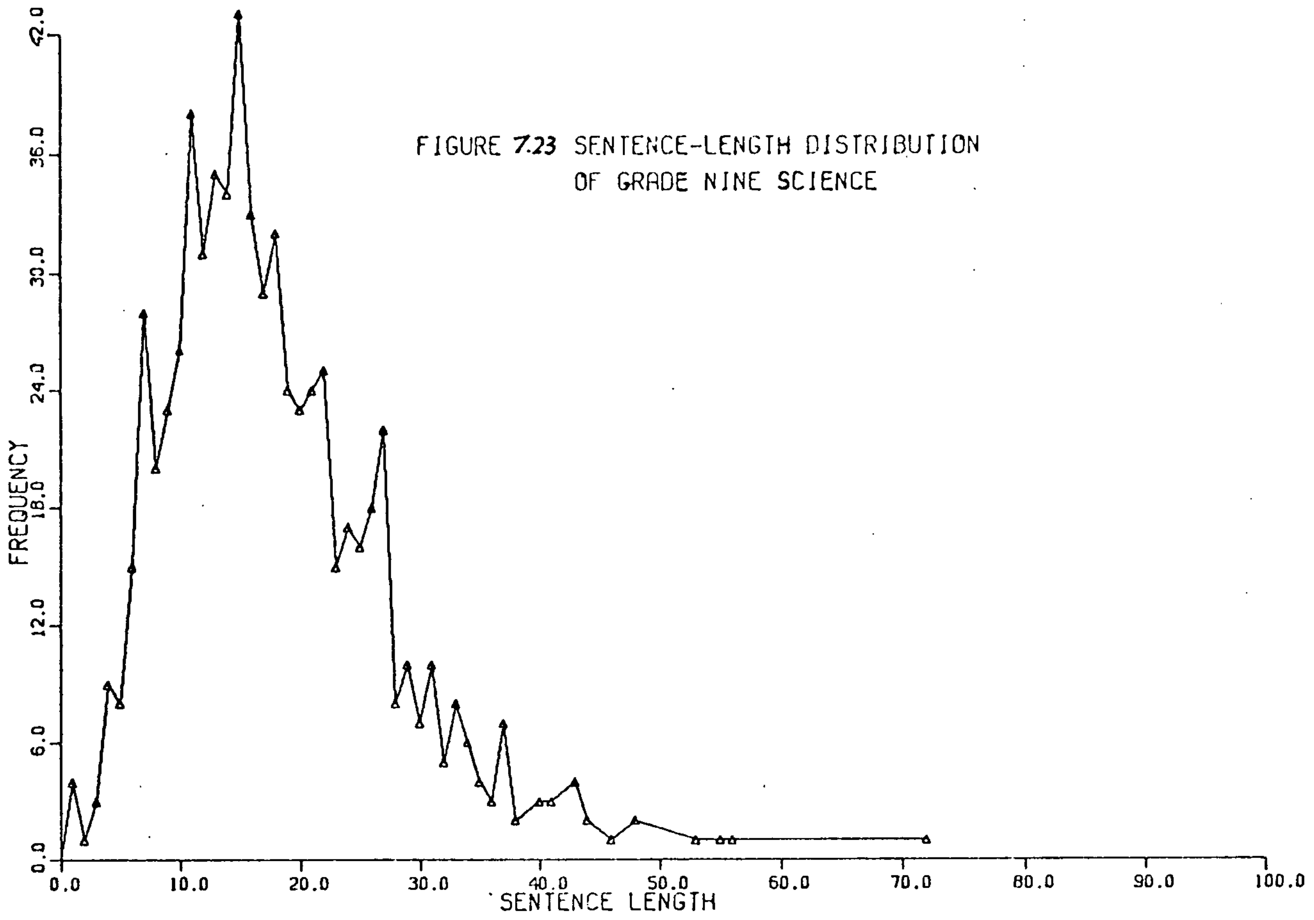


FIGURE 7.23 SENTENCE-LENGTH DISTRIBUTION
OF GRADE NINE SCIENCE
FREQUENCY
0.0
6.0
12.0
18.0
24.0
30.0
36.0
42.0
0.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0
SENTENCE LENGTH

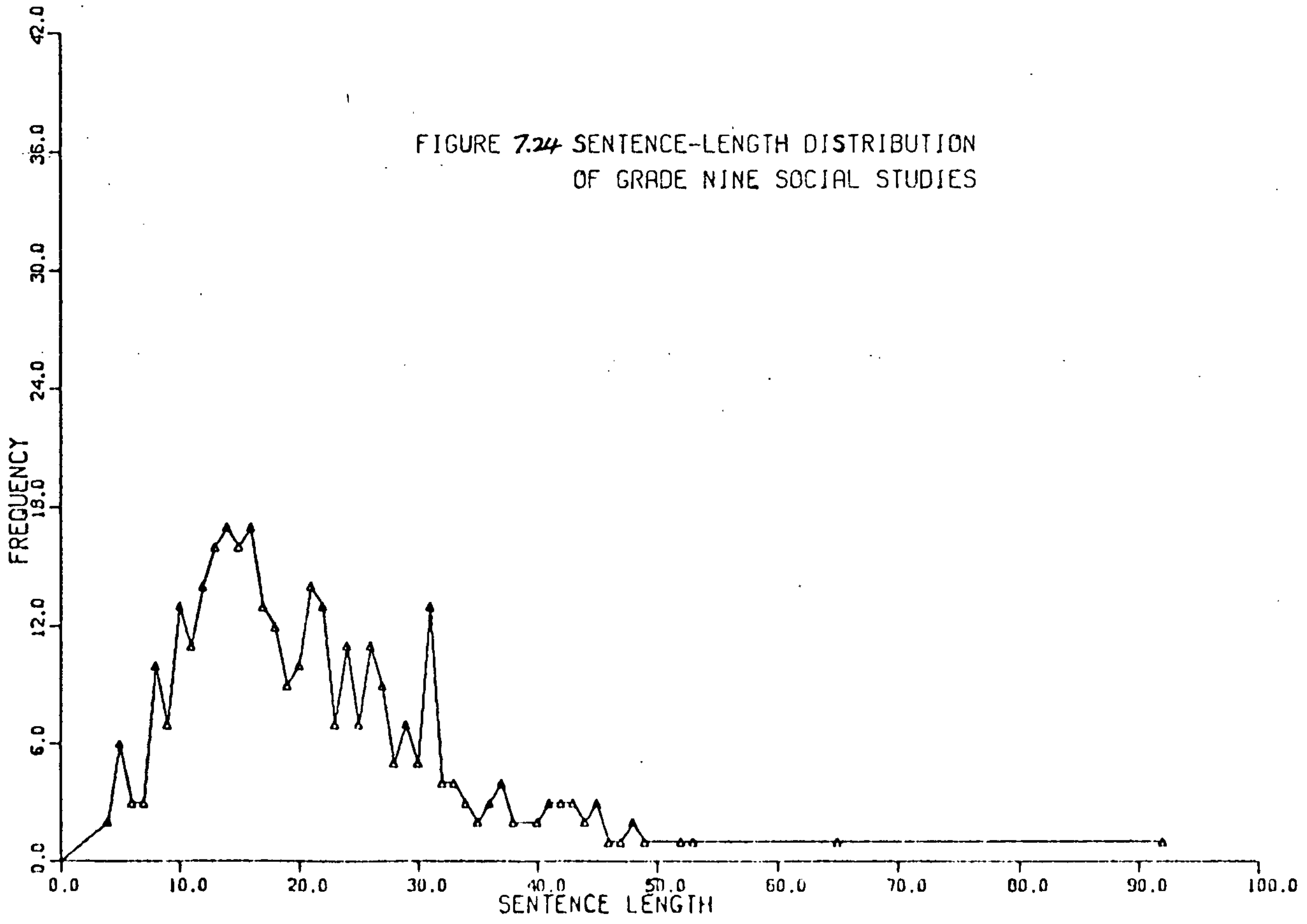


FIGURE 7.24 SENTENCE-LENGTH DISTRIBUTION
OF GRADE NINE SOCIAL STUDIES
FREQUENCY
0.0
6.0
12.0
18.0
24.0
30.0
36.0
42.0
0.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0
SENTENCE LENGTH

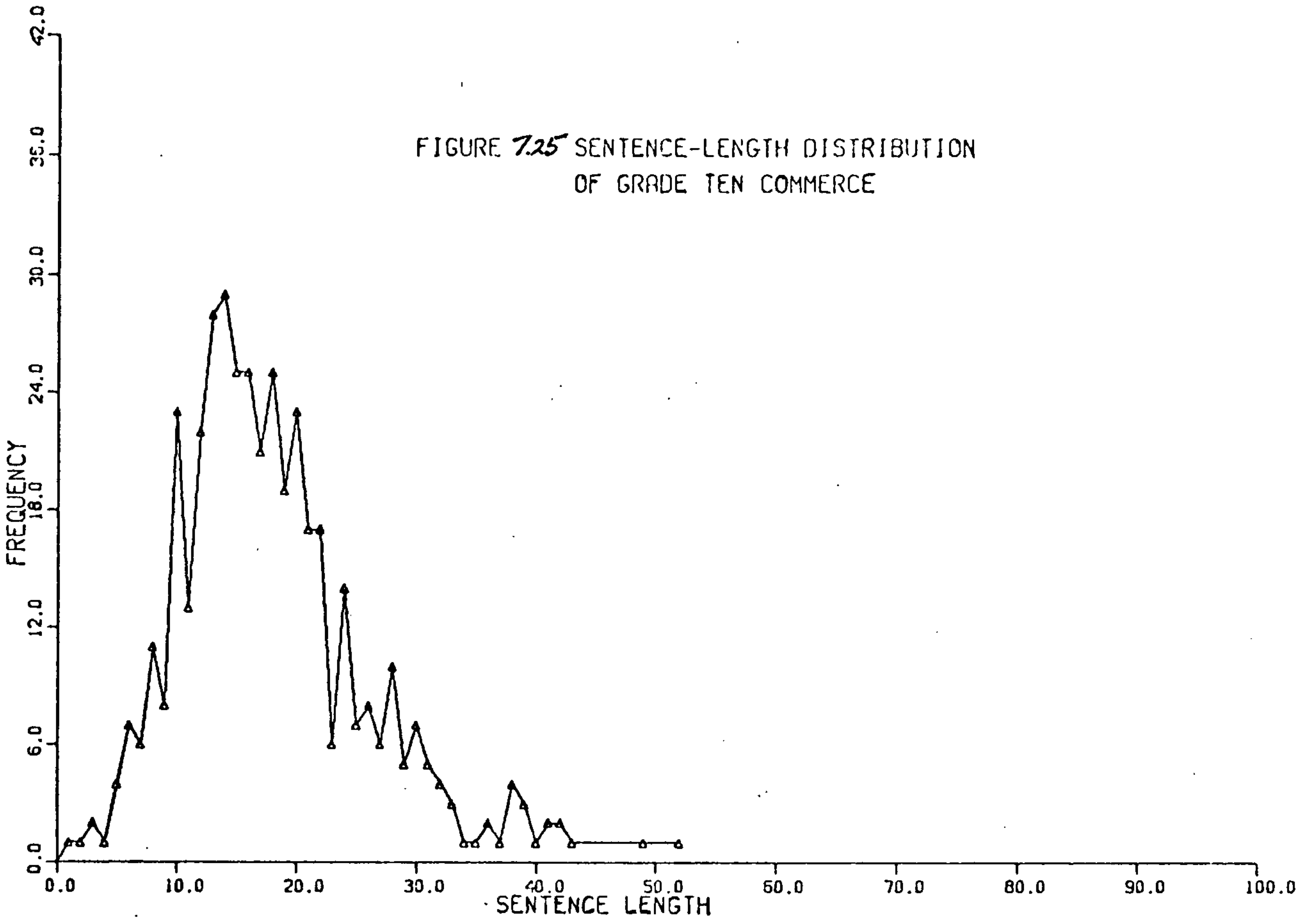


FIGURE 7.25 SENTENCE-LENGTH DISTRIBUTION
OF GRADE TEN COMMERCE
FREQUENCY
SENTENCE LENGTH
0.0
6.0
12.0
18.0
24.0
30.0
36.0
42.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0

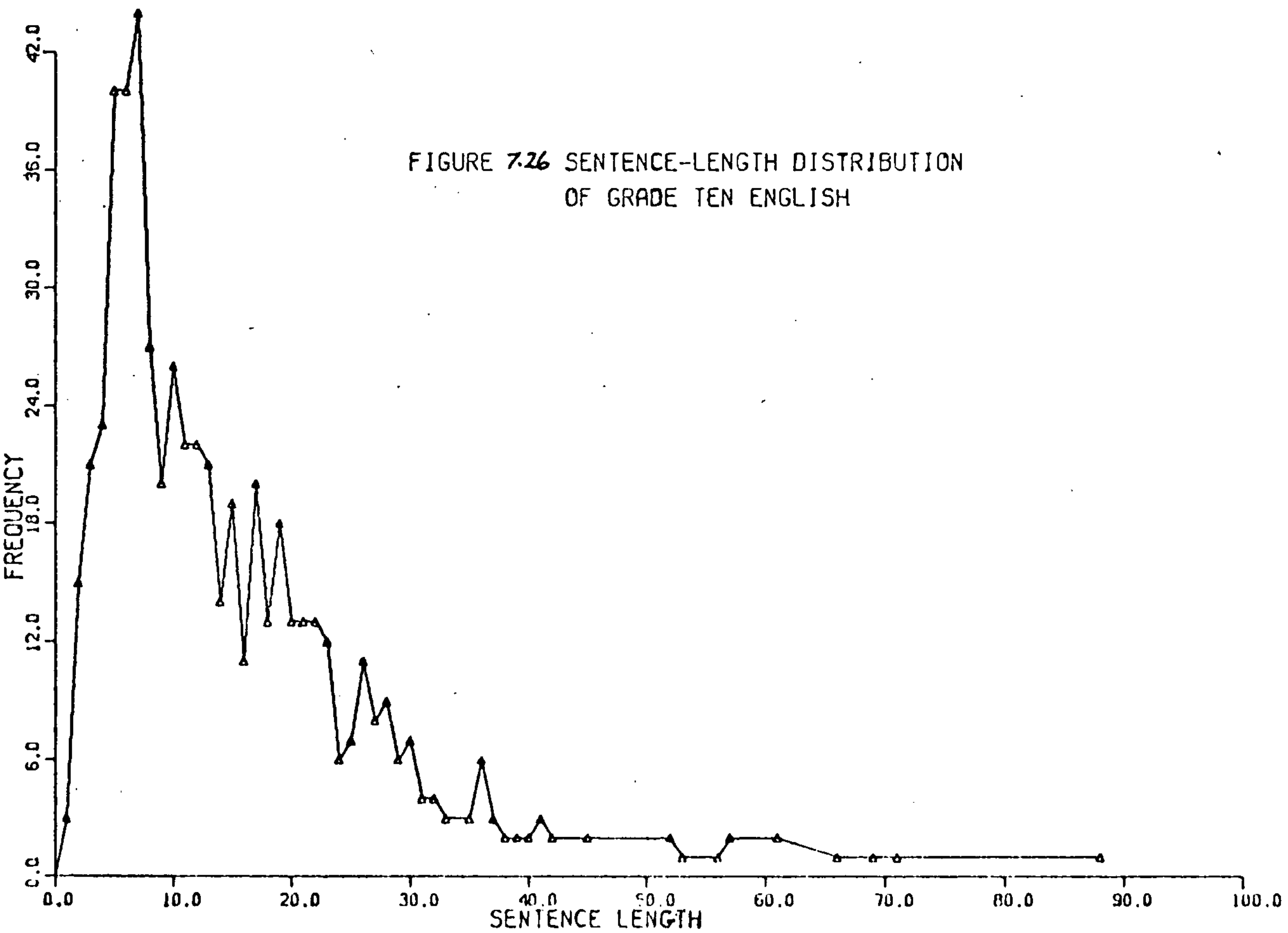


FIGURE 7.26 SENTENCE-LENGTH DISTRIBUTION
OF GRADE TEN ENGLISH
FREQUENCY
SENTENCE LENGTH
0.0
6.0
12.0
18.0
24.0
30.0
36.0
42.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0

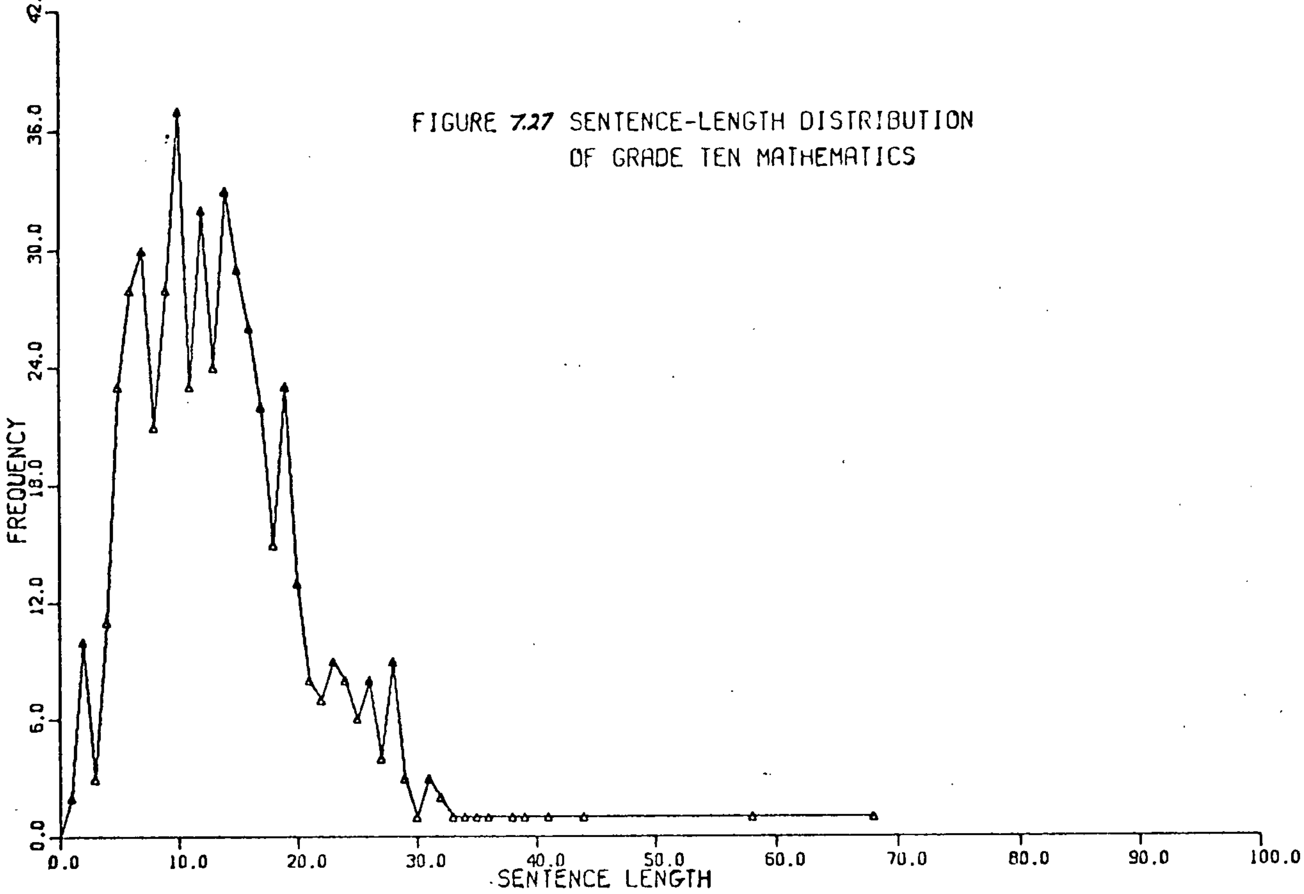


FIGURE 7.27 SENTENCE-LENGTH DISTRIBUTION
OF GRADE TEN MATHEMATICS
FREQUENCY
0.0
6.0
12.0
18.0
24.0
30.0
36.0
42.0
0.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0
SENTENCE LENGTH

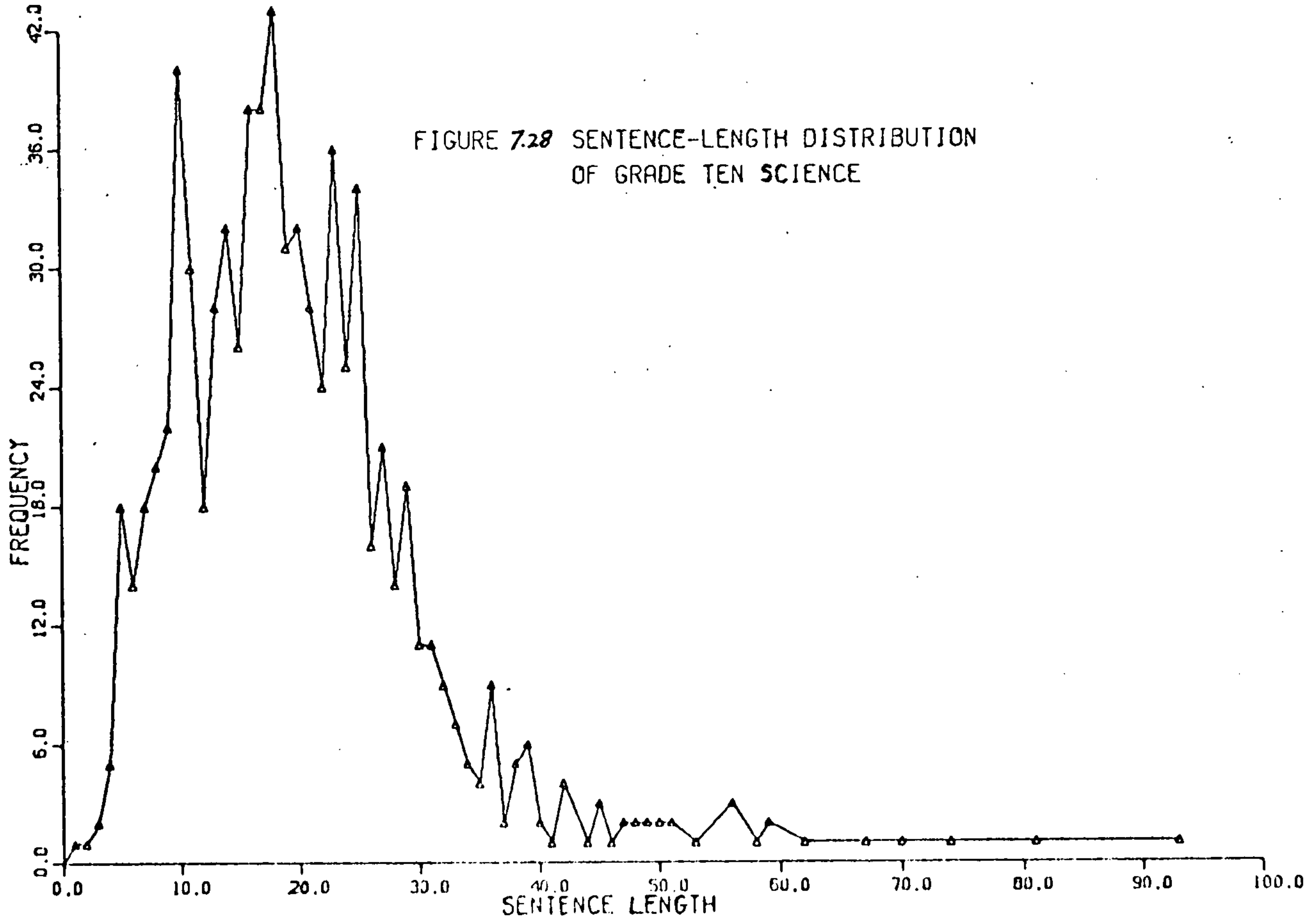


FIGURE 7.28 SENTENCE-LENGTH DISTRIBUTION
OF GRADE TEN SCIENCE
FREQUENCY
0.0
6.0
12.0
18.0
24.0
30.0
36.0
42.0
0.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0
SENTENCE LENGTH

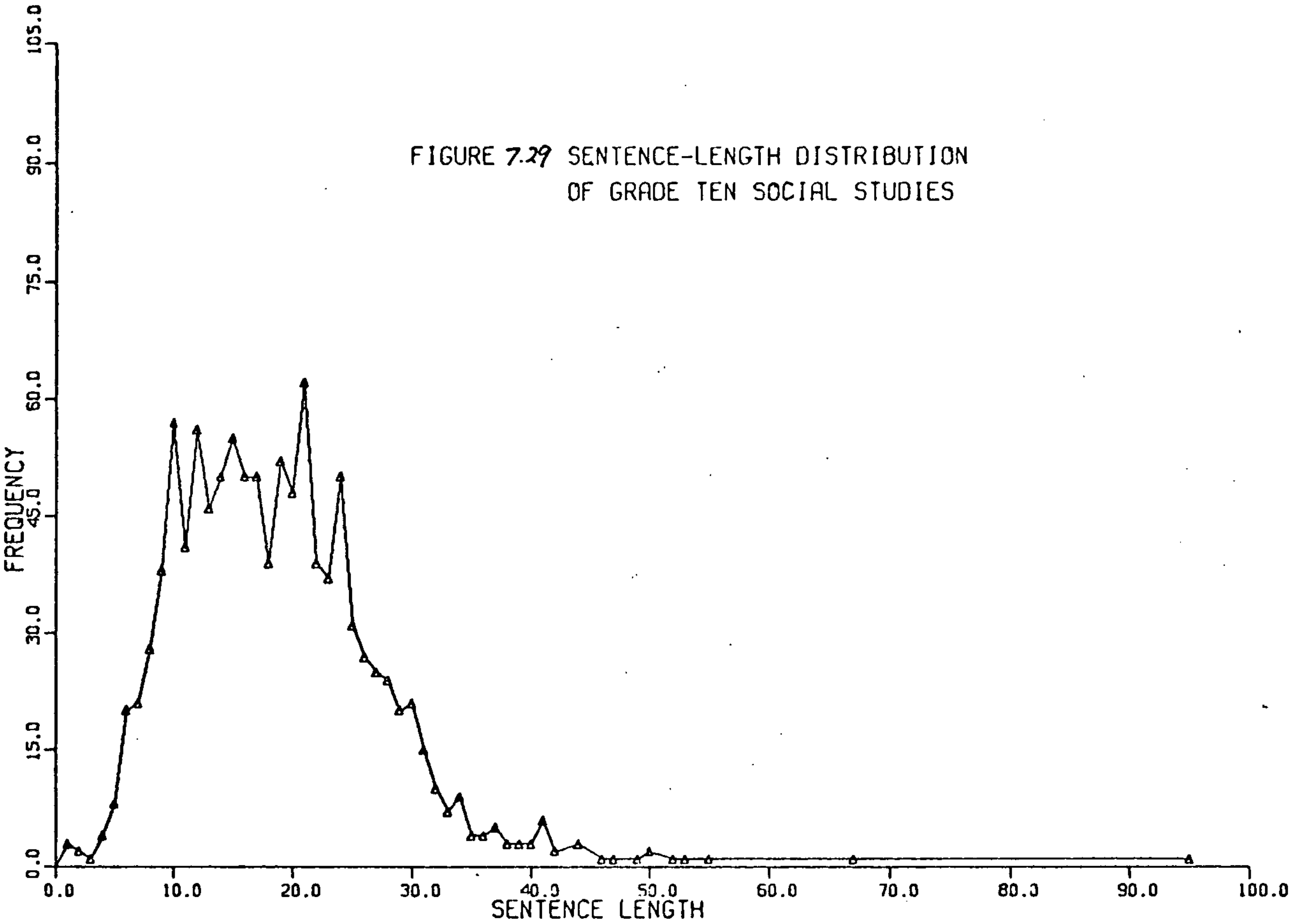


FIGURE 7.29 SENTENCE-LENGTH DISTRIBUTION OF GRADE TEN SOCIAL STUDIES

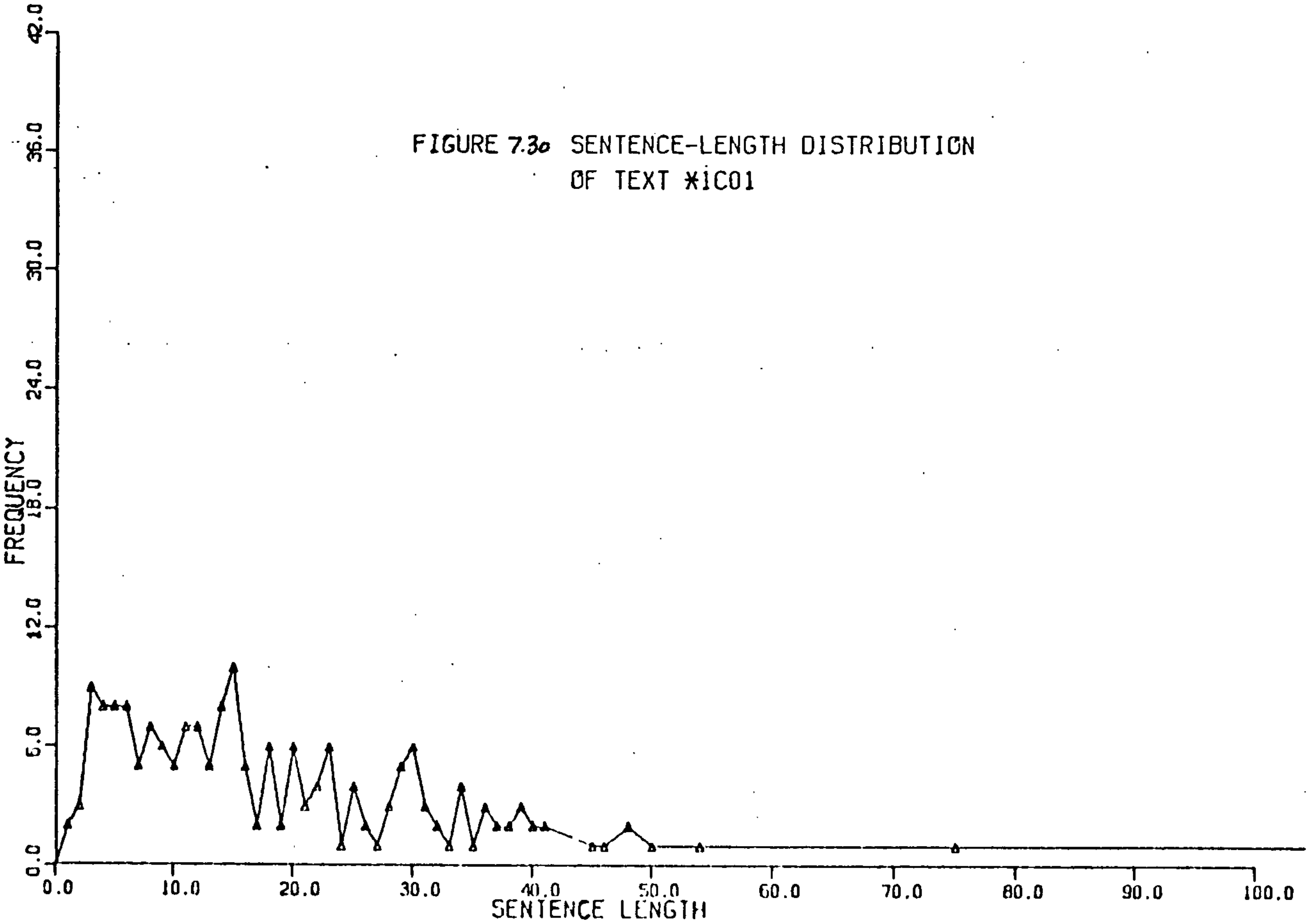


FIGURE 7.30 SENTENCE-LENGTH DISTRIBUTION OF TEXT *IC01

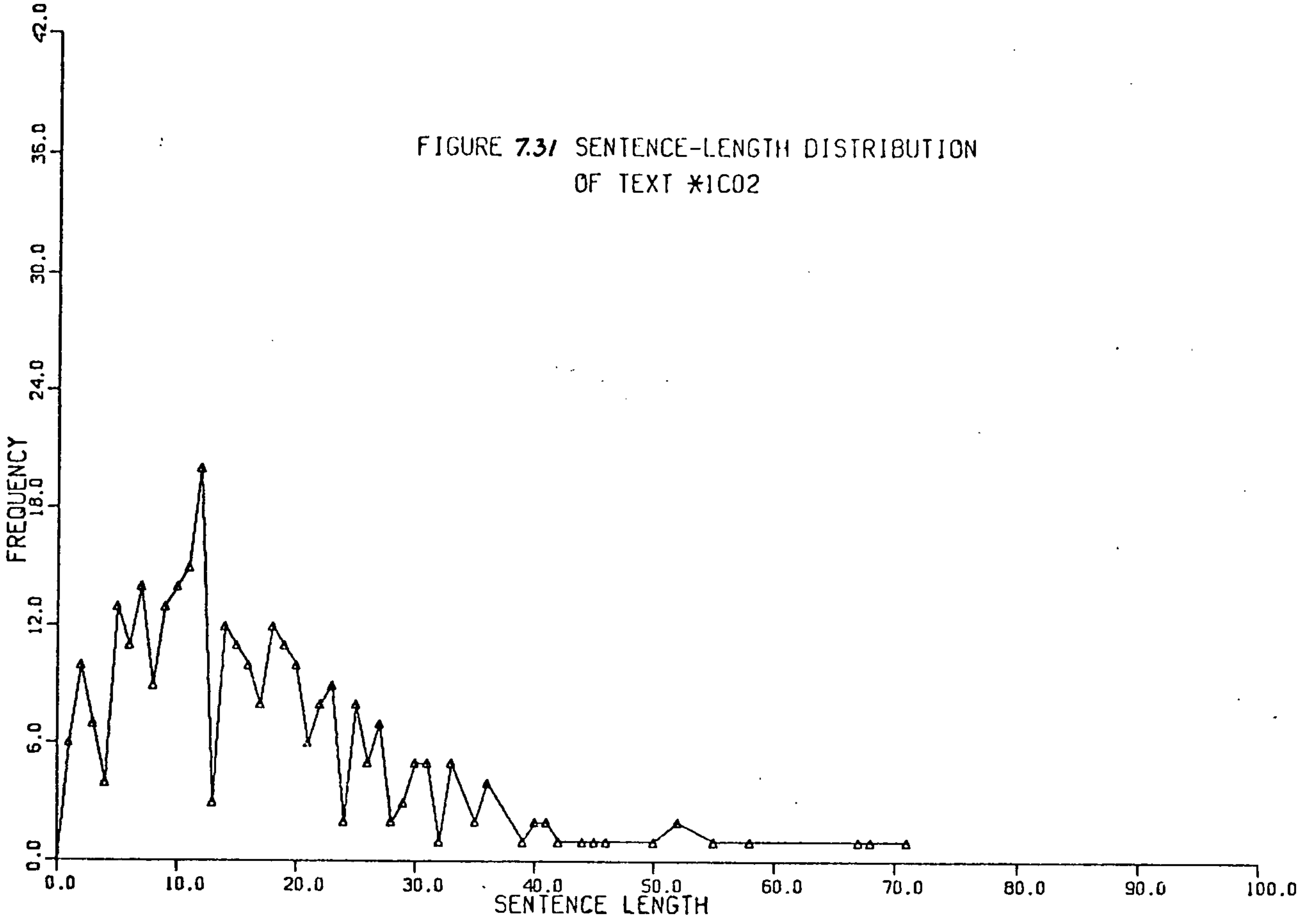


FIGURE 7.31 SENTENCE-LENGTH DISTRIBUTION OF TEXT *1C02

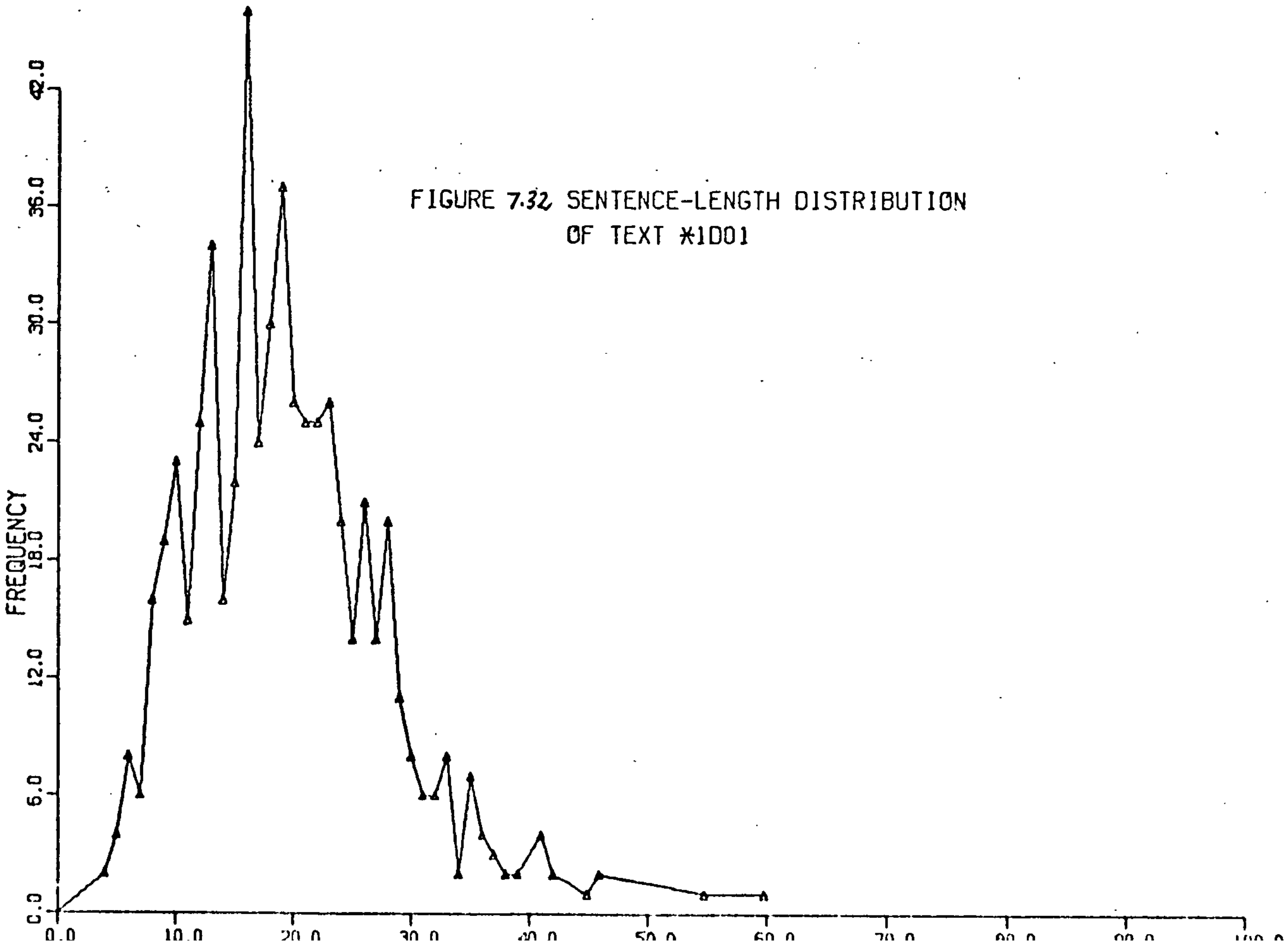


FIGURE 7.32 SENTENCE-LENGTH DISTRIBUTION OF TEXT *1D01

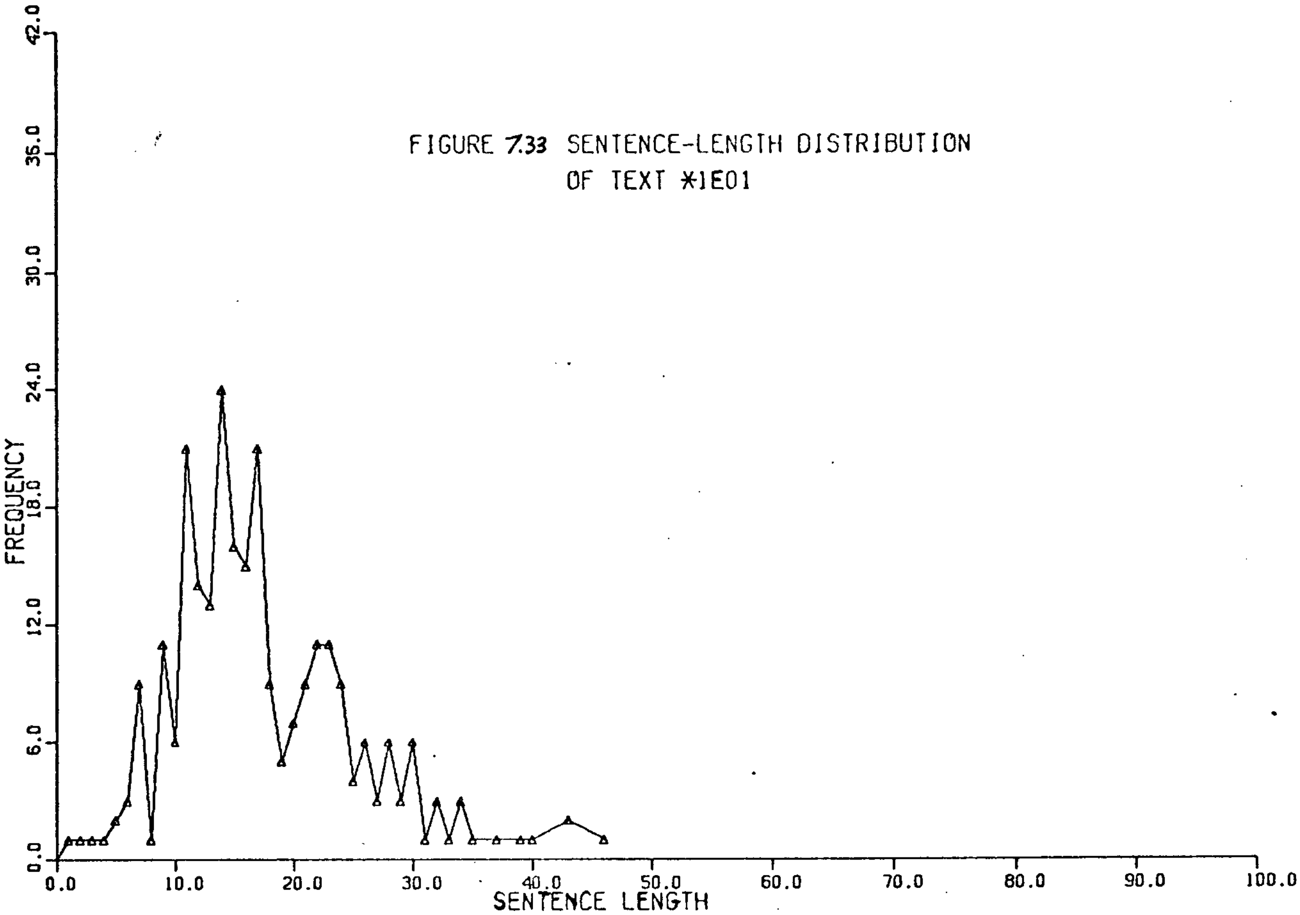


FIGURE 7.33 SENTENCE-LENGTH DISTRIBUTION OF TEXT *1E01

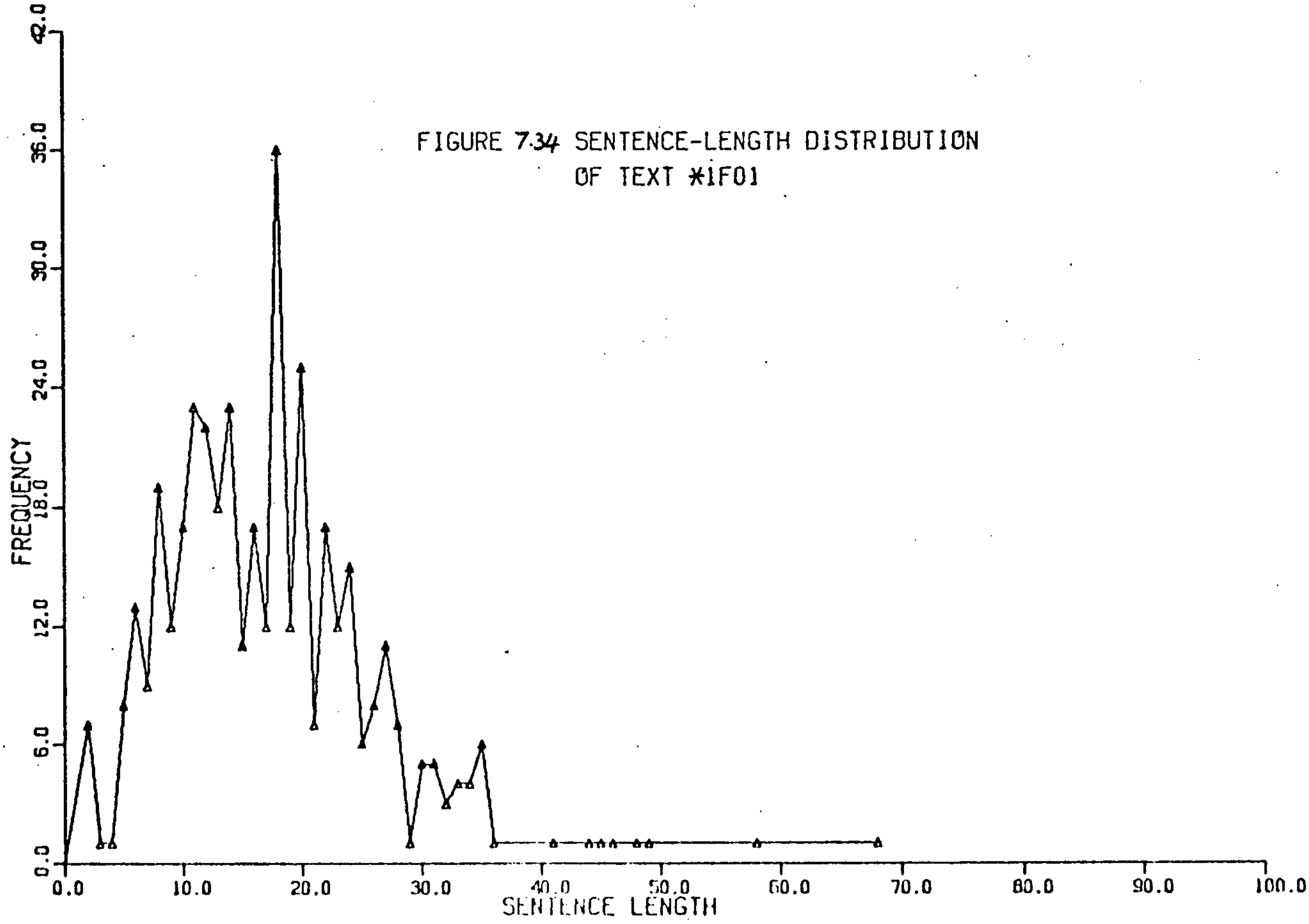


FIGURE 7.34 SENTENCE-LENGTH DISTRIBUTION OF TEXT *1F01

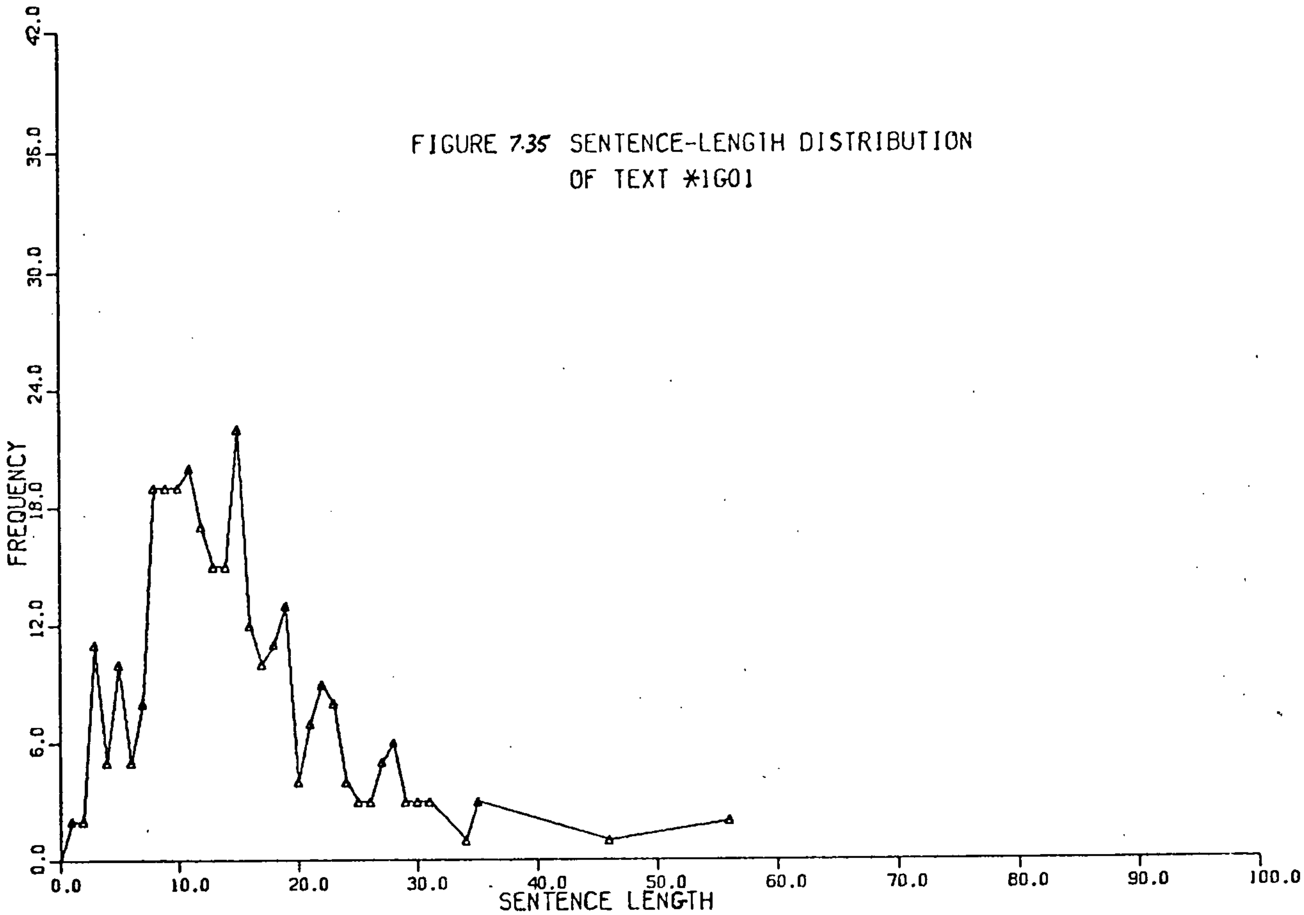


FIGURE 7.35 SENTENCE-LENGTH DISTRIBUTION
OF TEXT *1G01
FREQUENCY
SENTENCE LENGTH
0.0
6.0
12.0
18.0
24.0
30.0
36.0
42.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0

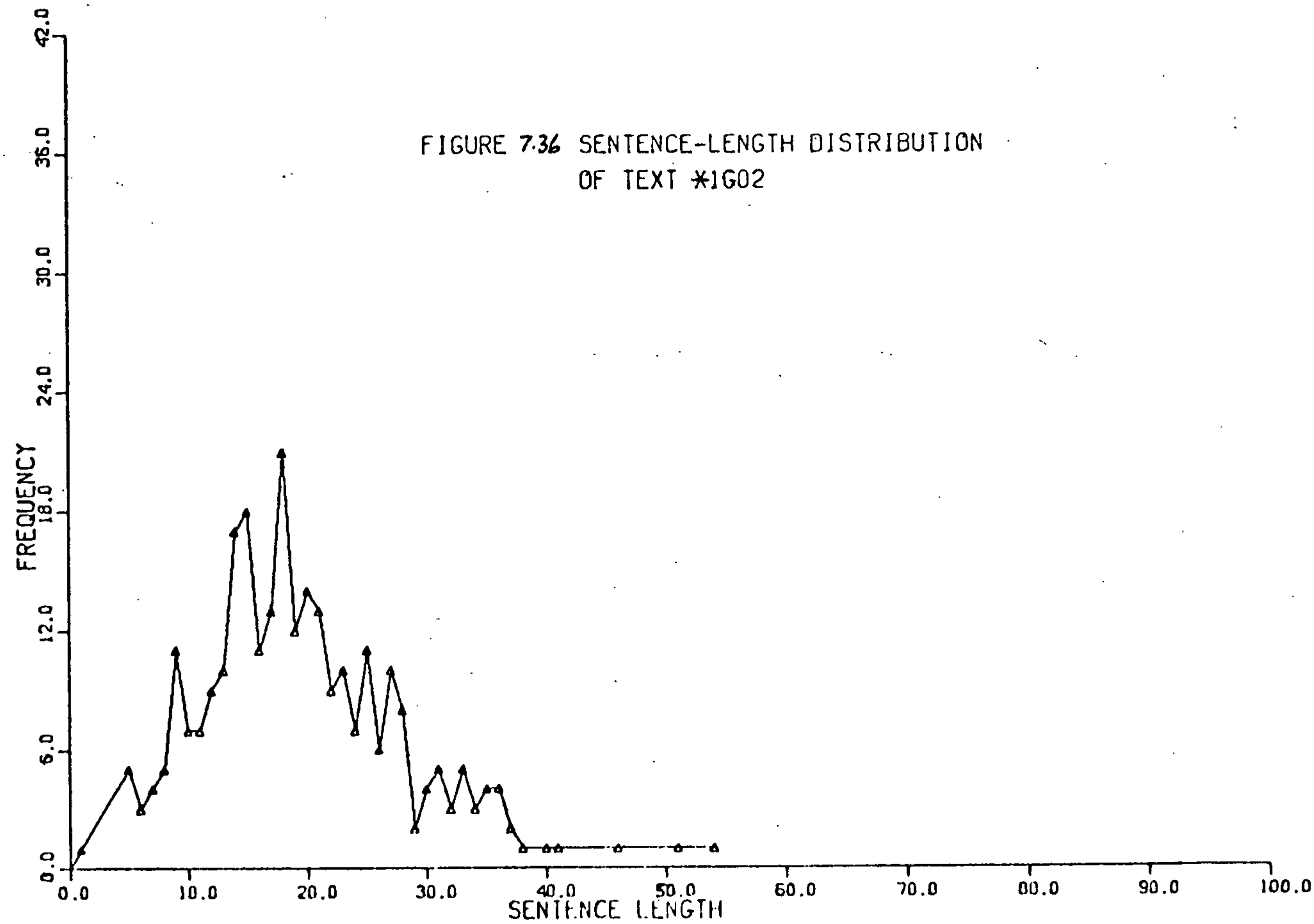


FIGURE 7.36 SENTENCE-LENGTH DISTRIBUTION
OF TEXT *1G02
FREQUENCY
SENTENCE LENGTH
0.0
6.0
12.0
18.0
24.0
30.0
36.0
42.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0

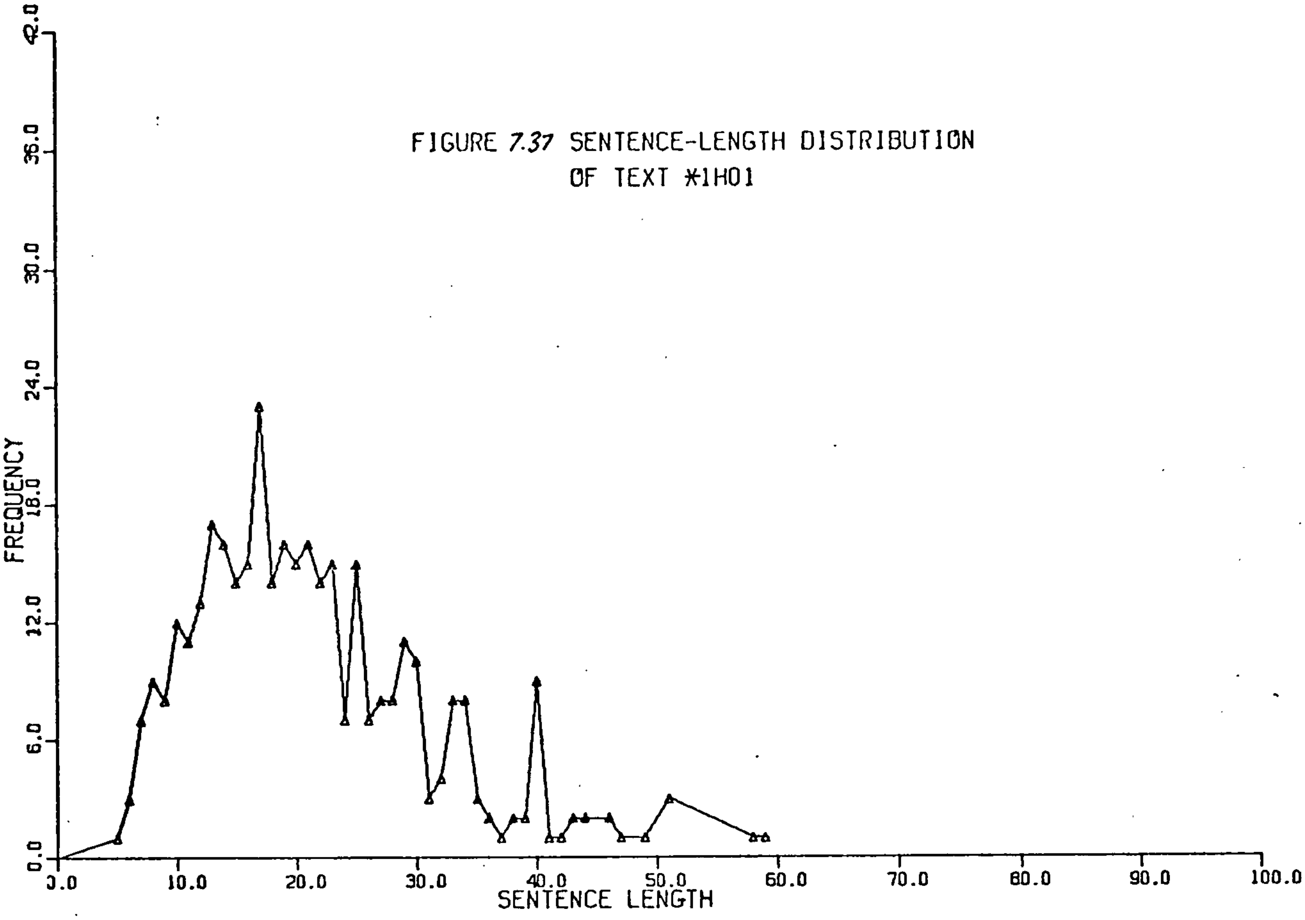


FIGURE 7.37 SENTENCE-LENGTH DISTRIBUTION OF TEXT *1H01

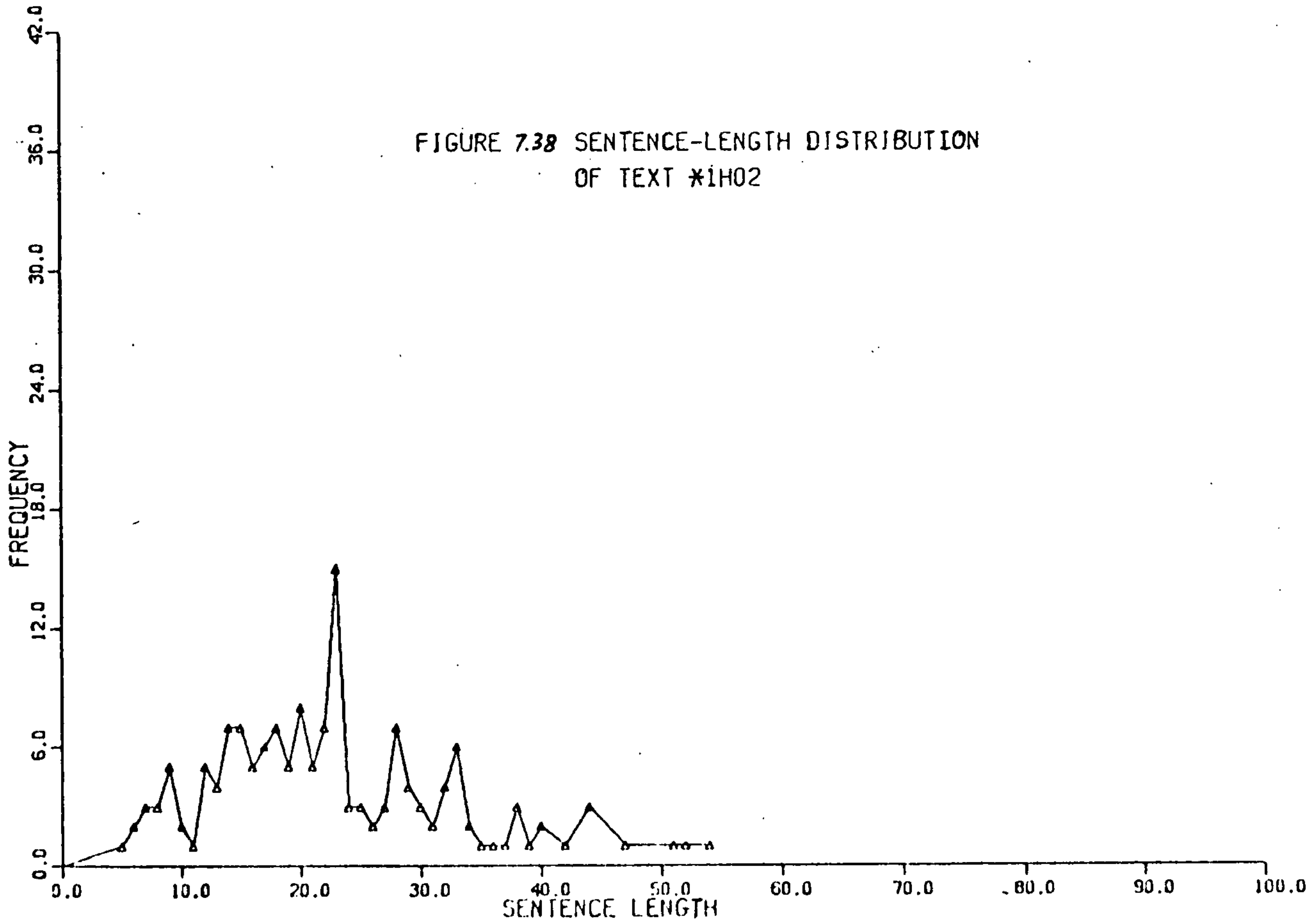


FIGURE 7.38 SENTENCE-LENGTH DISTRIBUTION OF TEXT *1H02

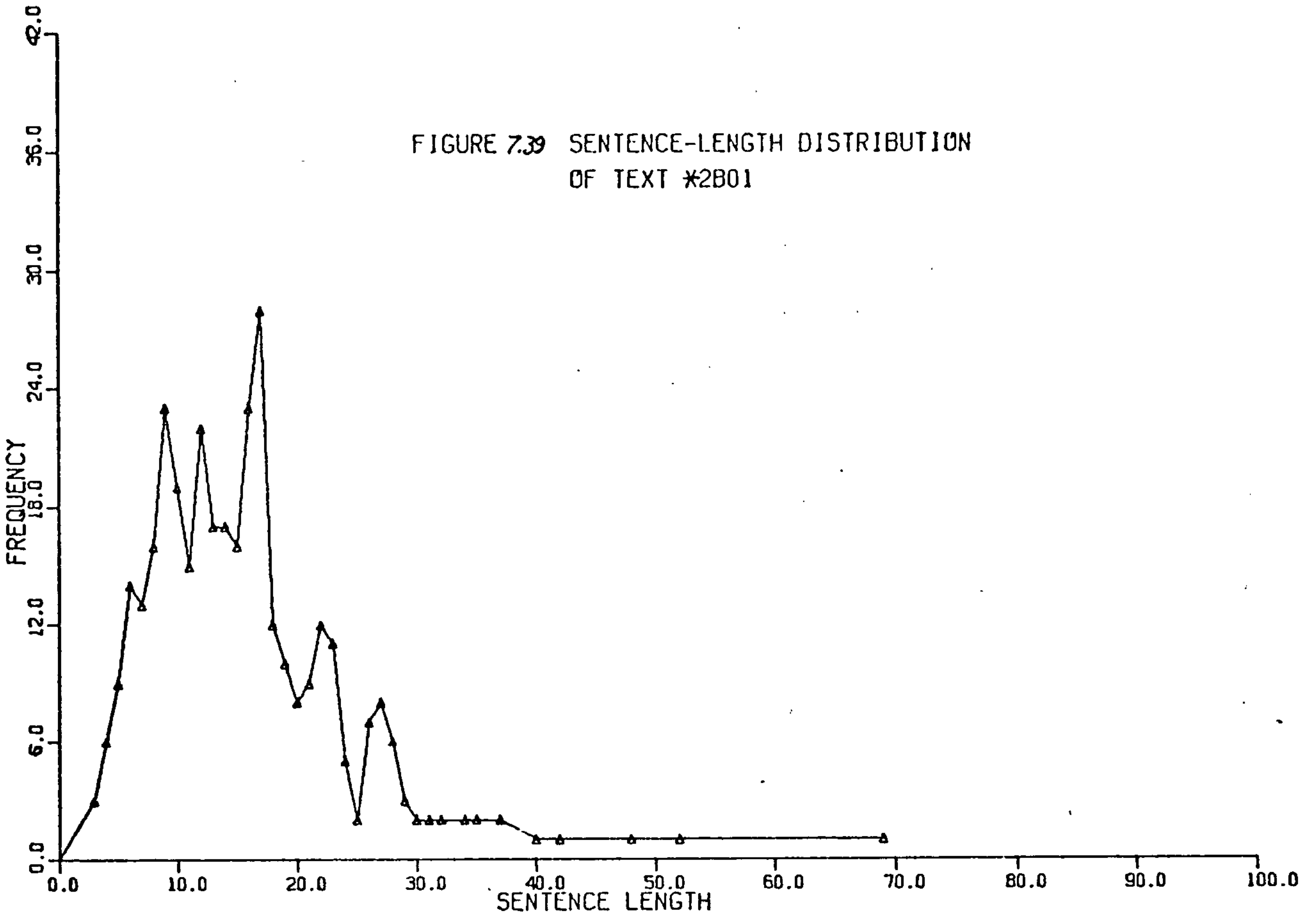


FIGURE 7.39 SENTENCE-LENGTH DISTRIBUTION
OF TEXT *2B01
FREQUENCY
SENTENCE LENGTH
0.0
6.0
12.0
18.0
24.0
30.0
36.0
42.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0

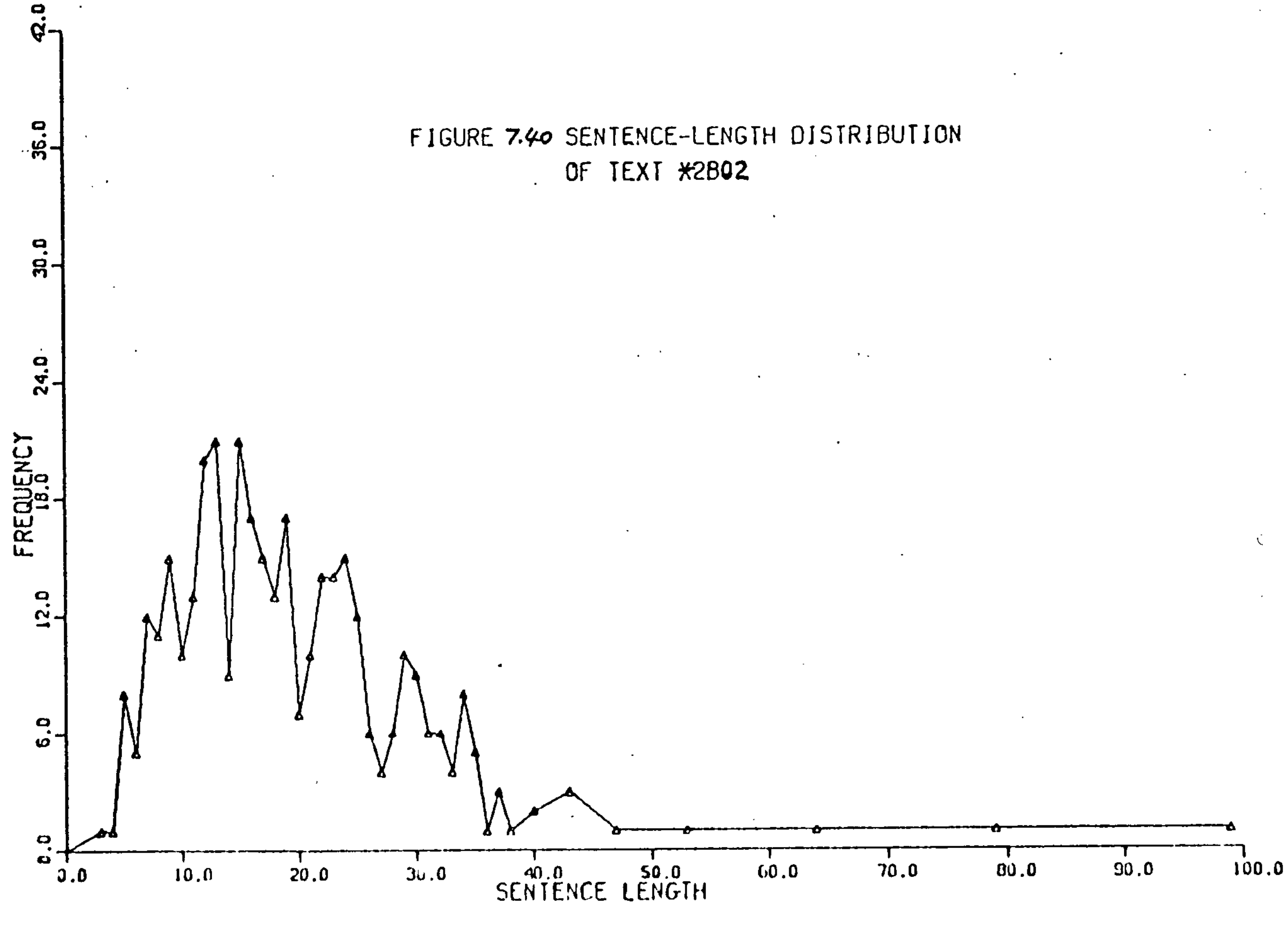


FIGURE 7.40 SENTENCE-LENGTH DISTRIBUTION
OF TEXT *2B02
FREQUENCY
SENTENCE LENGTH
0.0
6.0
12.0
18.0
24.0
30.0
36.0
42.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0

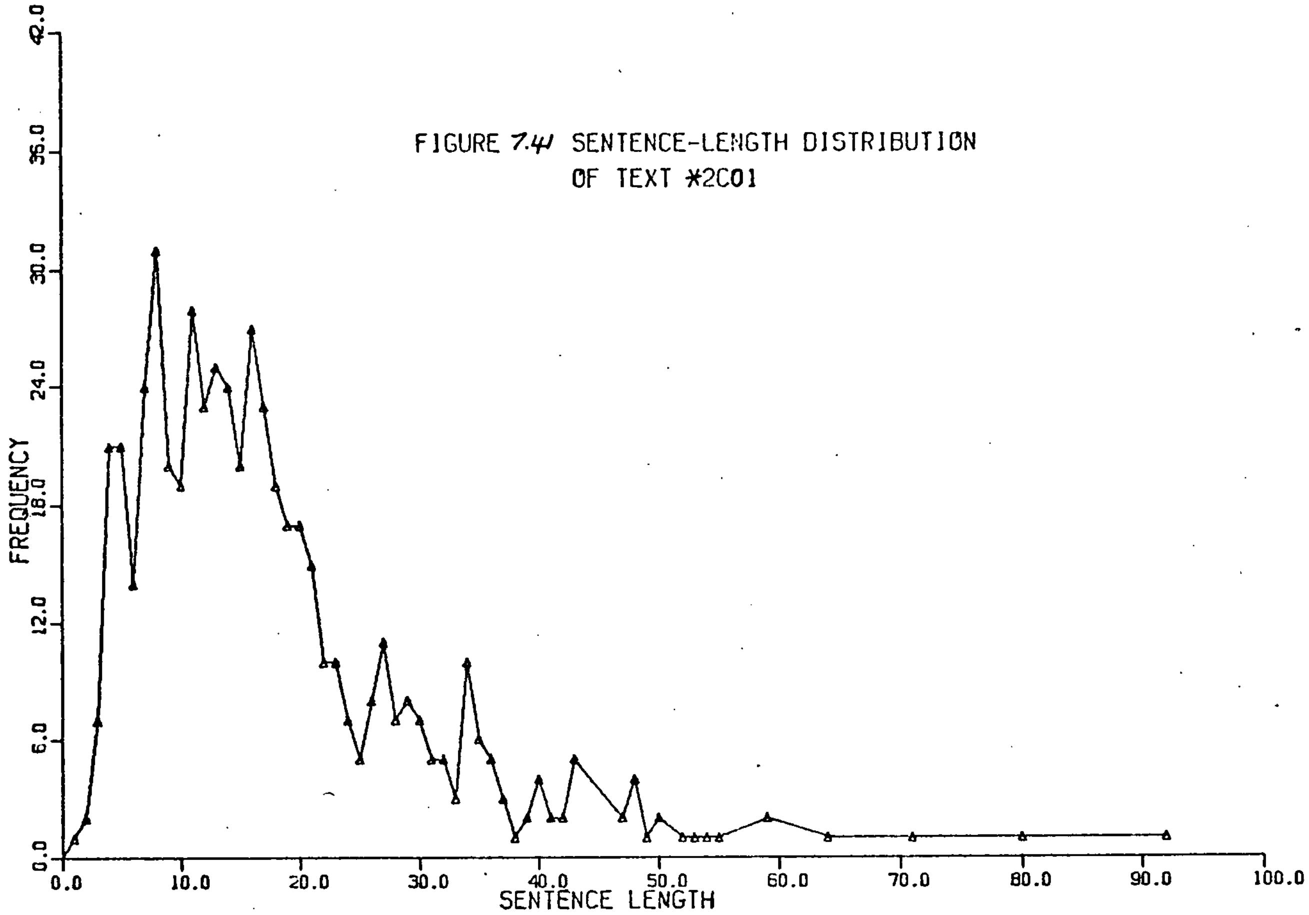


FIGURE 7.41 SENTENCE-LENGTH DISTRIBUTION
OF TEXT *2C01
FREQUENCY
0.0
6.0
12.0
18.0
24.0
30.0
36.0
42.0
0.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0
SENTENCE LENGTH

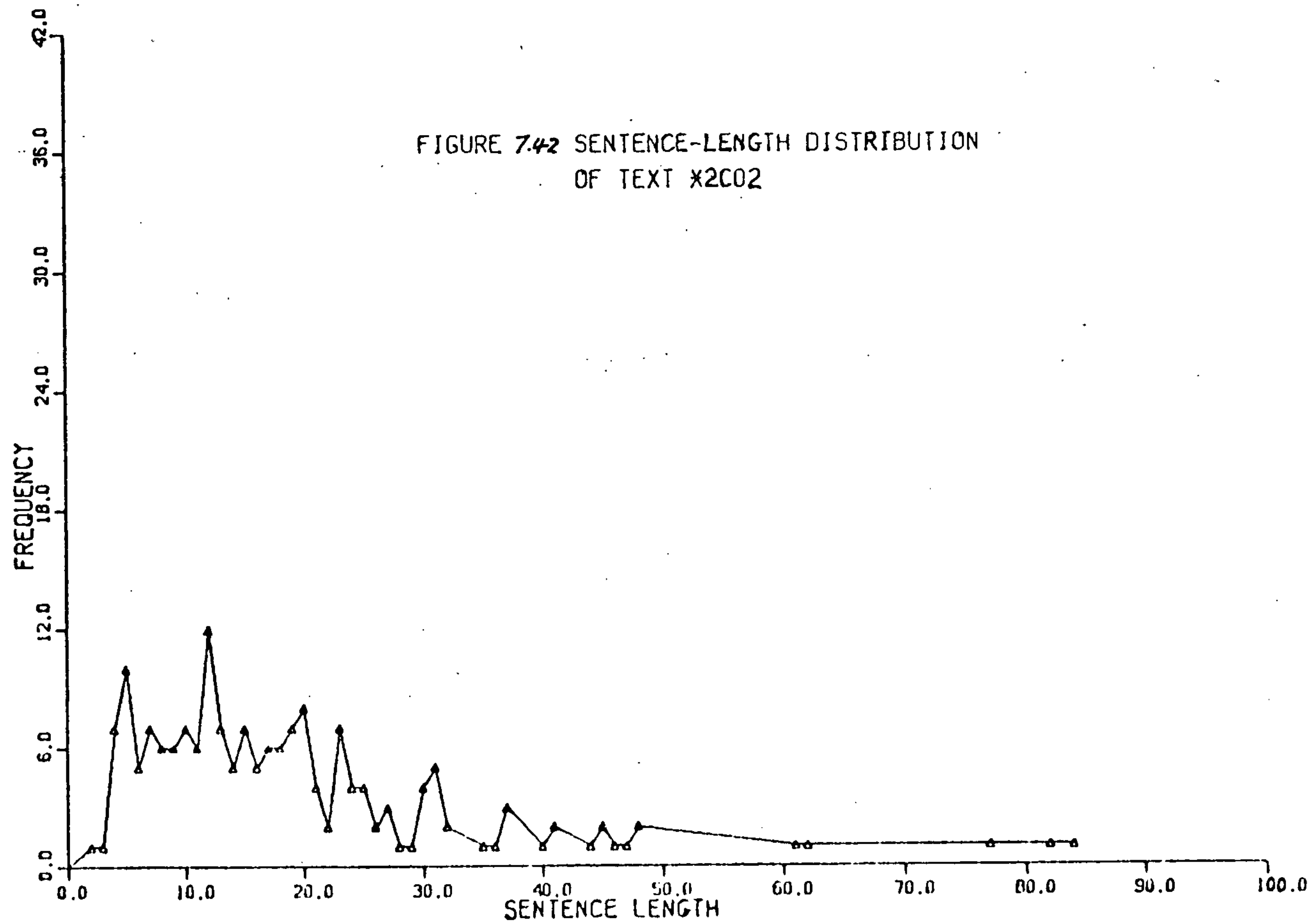


FIGURE 7.42 SENTENCE-LENGTH DISTRIBUTION
OF TEXT *2C02
FREQUENCY
0.0
6.0
12.0
18.0
24.0
30.0
36.0
42.0
0.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0
SENTENCE LENGTH

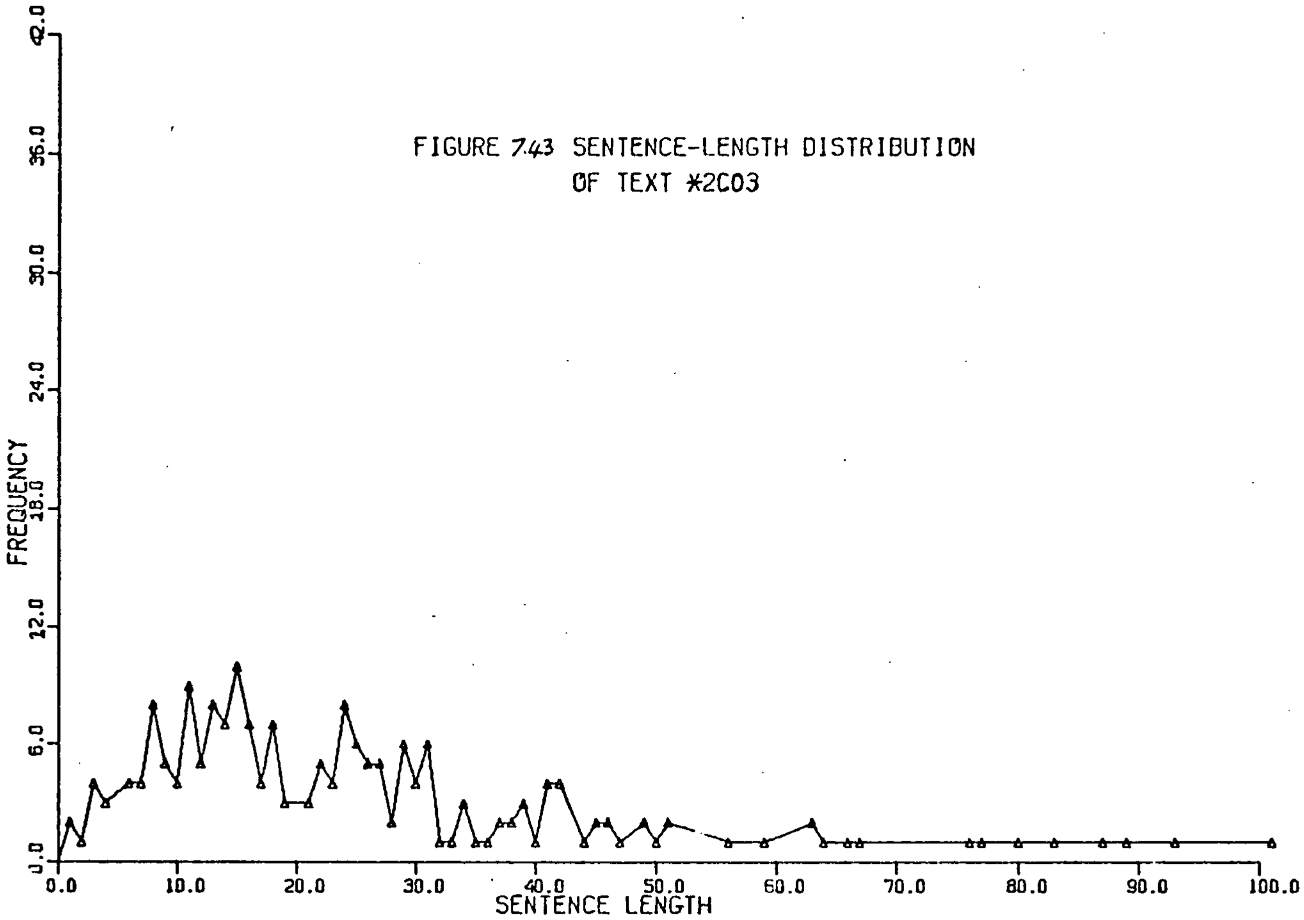


FIGURE 7.43 SENTENCE-LENGTH DISTRIBUTION
OF TEXT *2C03
FREQUENCY
SENTENCE LENGTH
0.0
6.0
12.0
18.0
24.0
30.0
36.0
42.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0

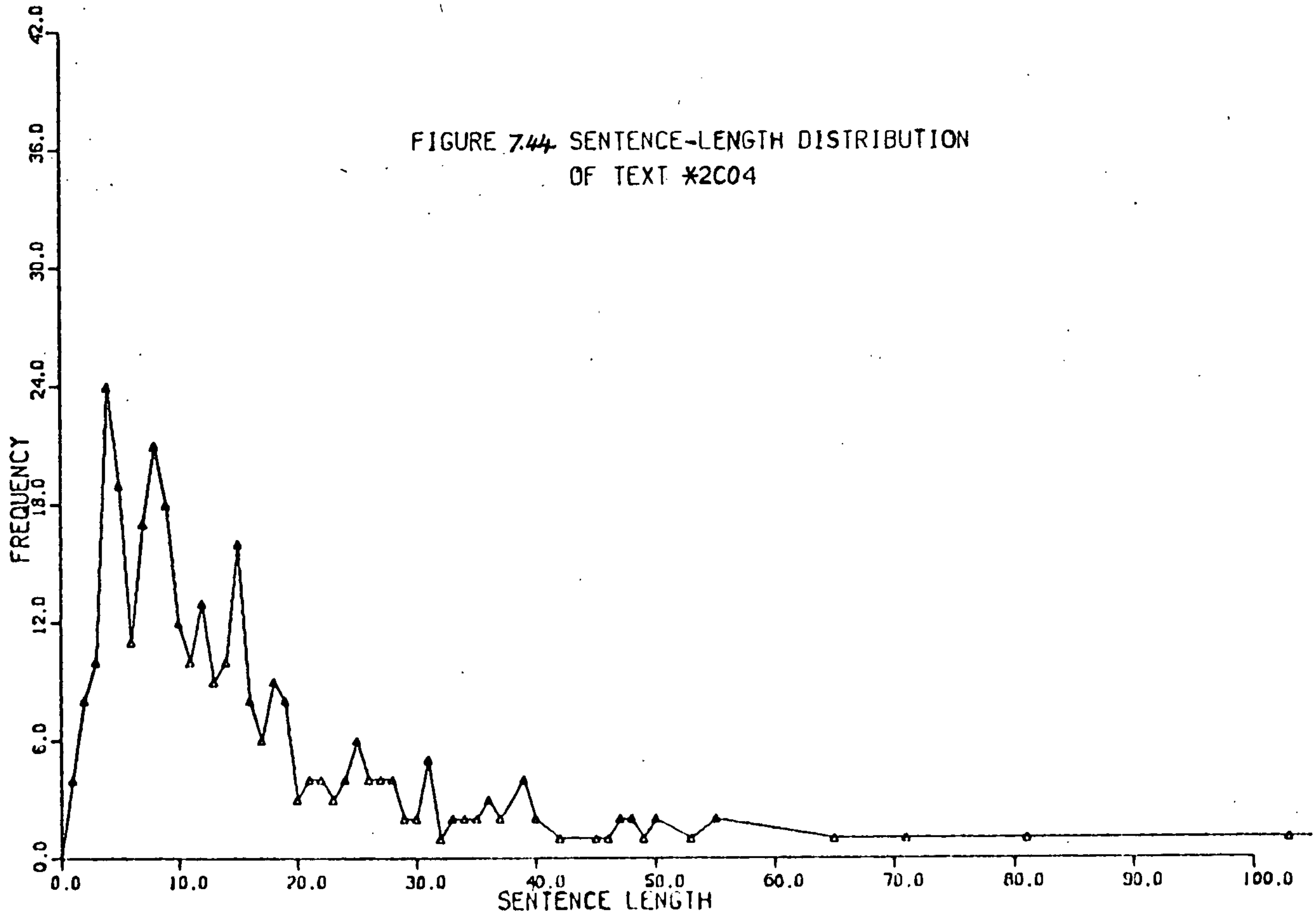


FIGURE 7.44 SENTENCE-LENGTH DISTRIBUTION
OF TEXT *2C04
FREQUENCY
SENTENCE LENGTH
0.0
6.0
12.0
18.0
24.0
30.0
36.0
42.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0

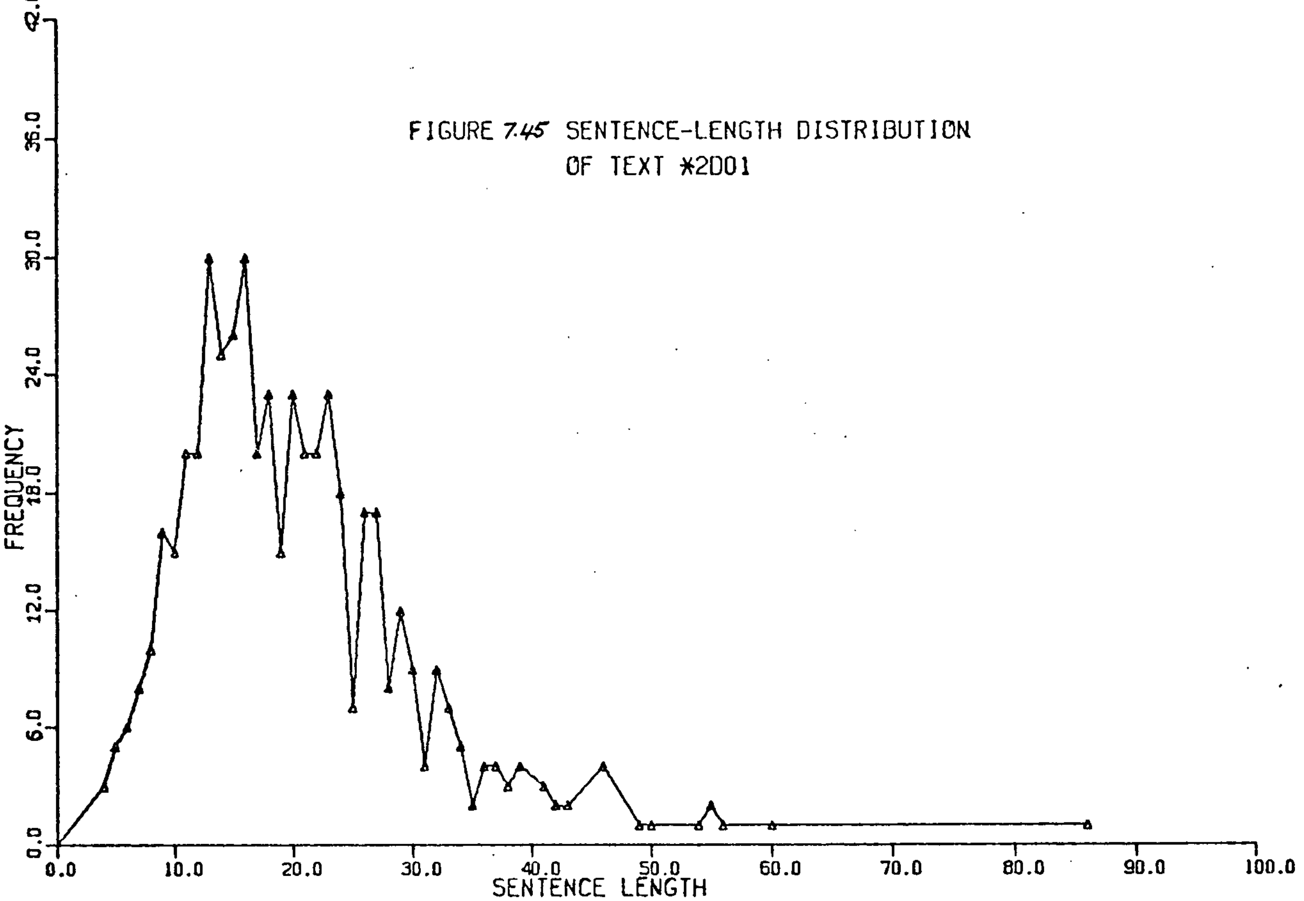


FIGURE 7.45 SENTENCE-LENGTH DISTRIBUTION
OF TEXT *2D01
FREQUENCY
SENTENCE LENGTH
0.0
6.0
12.0
18.0
24.0
30.0
36.0
42.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0

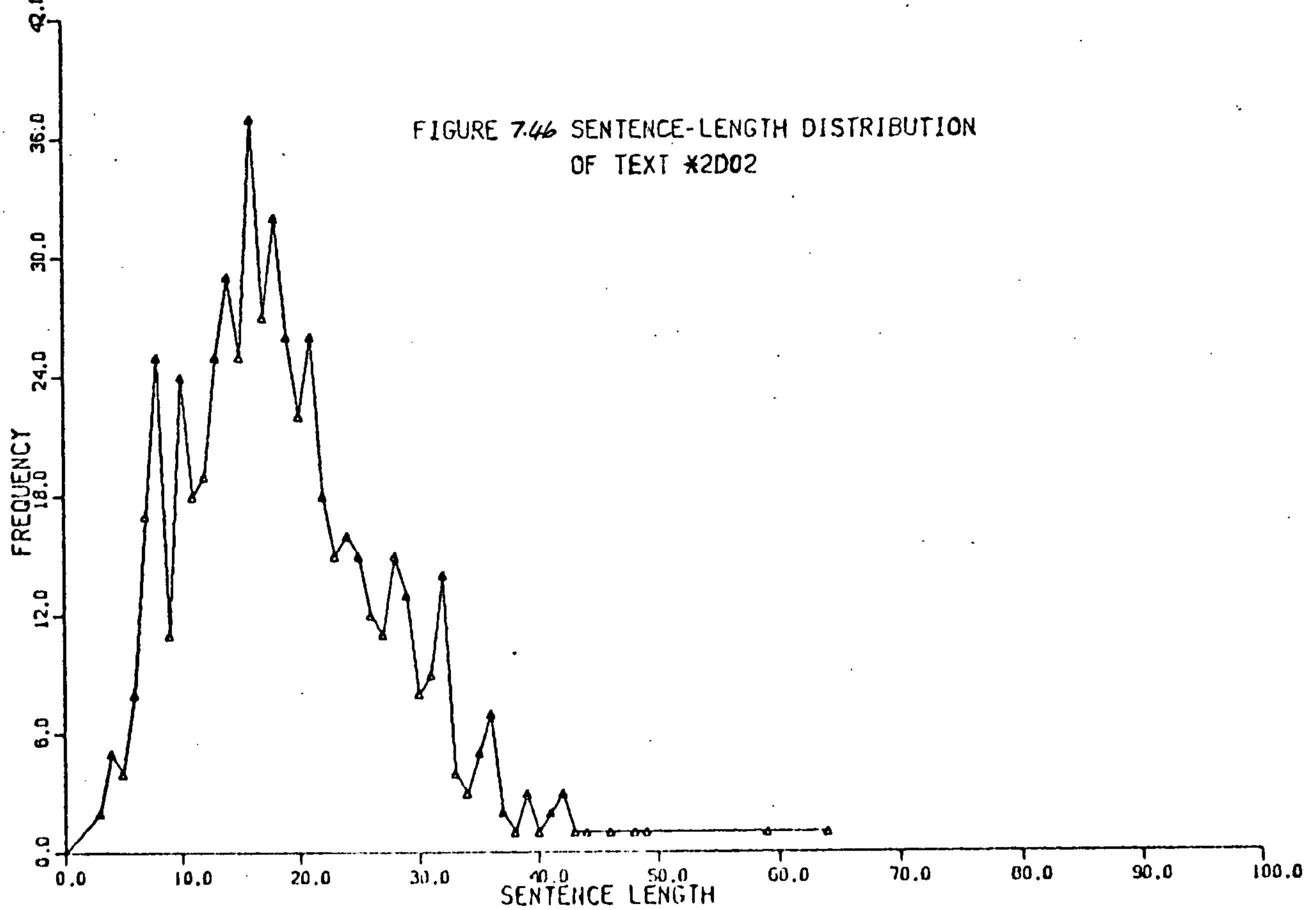


FIGURE 7.46 SENTENCE-LENGTH DISTRIBUTION
OF TEXT *2D02
FREQUENCY
SENTENCE LENGTH
0.0
6.0
12.0
18.0
24.0
30.0
36.0
42.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0

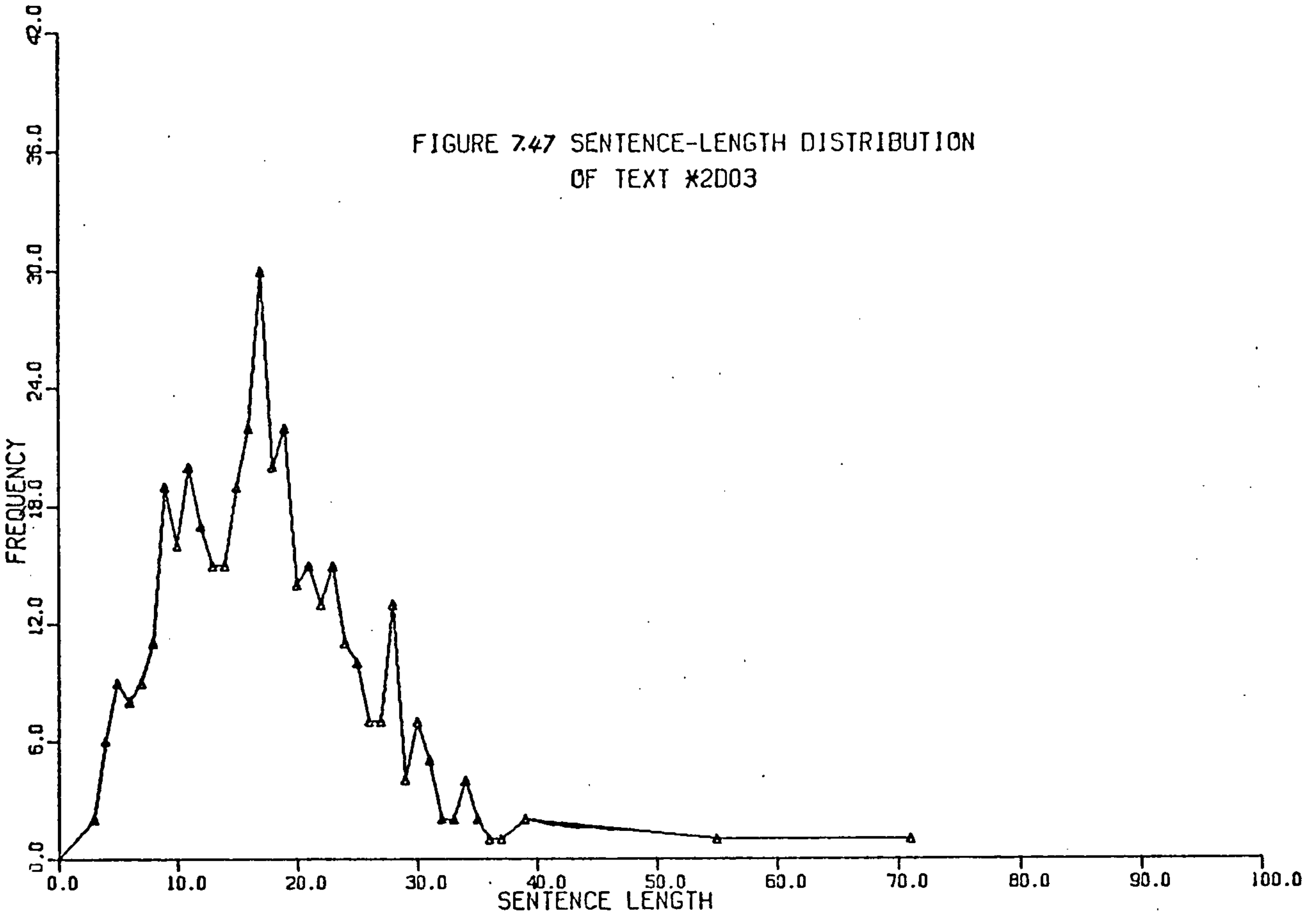


FIGURE 7.47 SENTENCE-LENGTH DISTRIBUTION
OF TEXT X2D03
FREQUENCY
SENTENCE LENGTH
0.0
6.0
12.0
18.0
24.0
30.0
36.0
42.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0

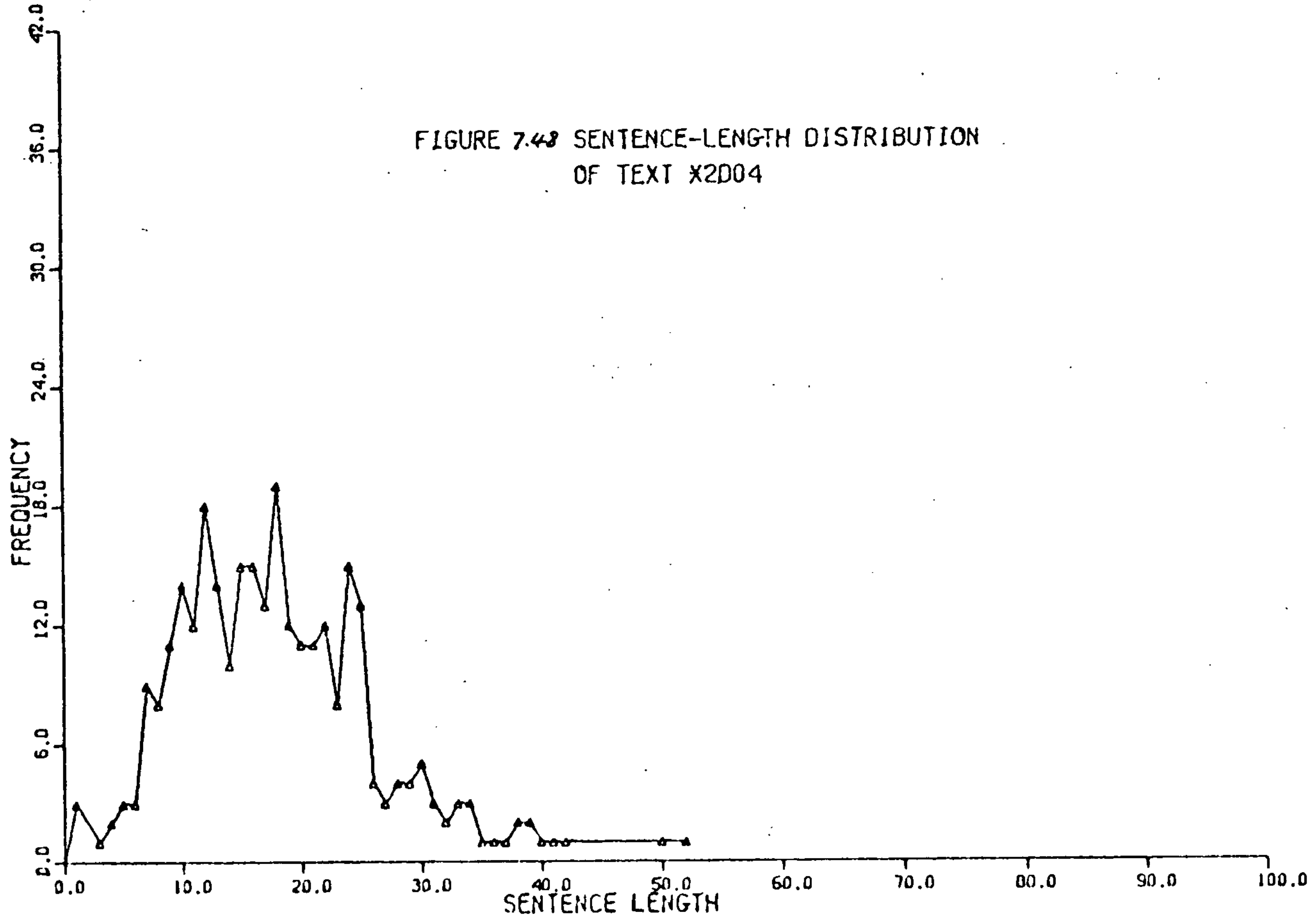


FIGURE 7.48 SENTENCE-LENGTH DISTRIBUTION
OF TEXT X2D04
FREQUENCY
SENTENCE LENGTH
0.0
6.0
12.0
18.0
24.0
30.0
36.0
42.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0

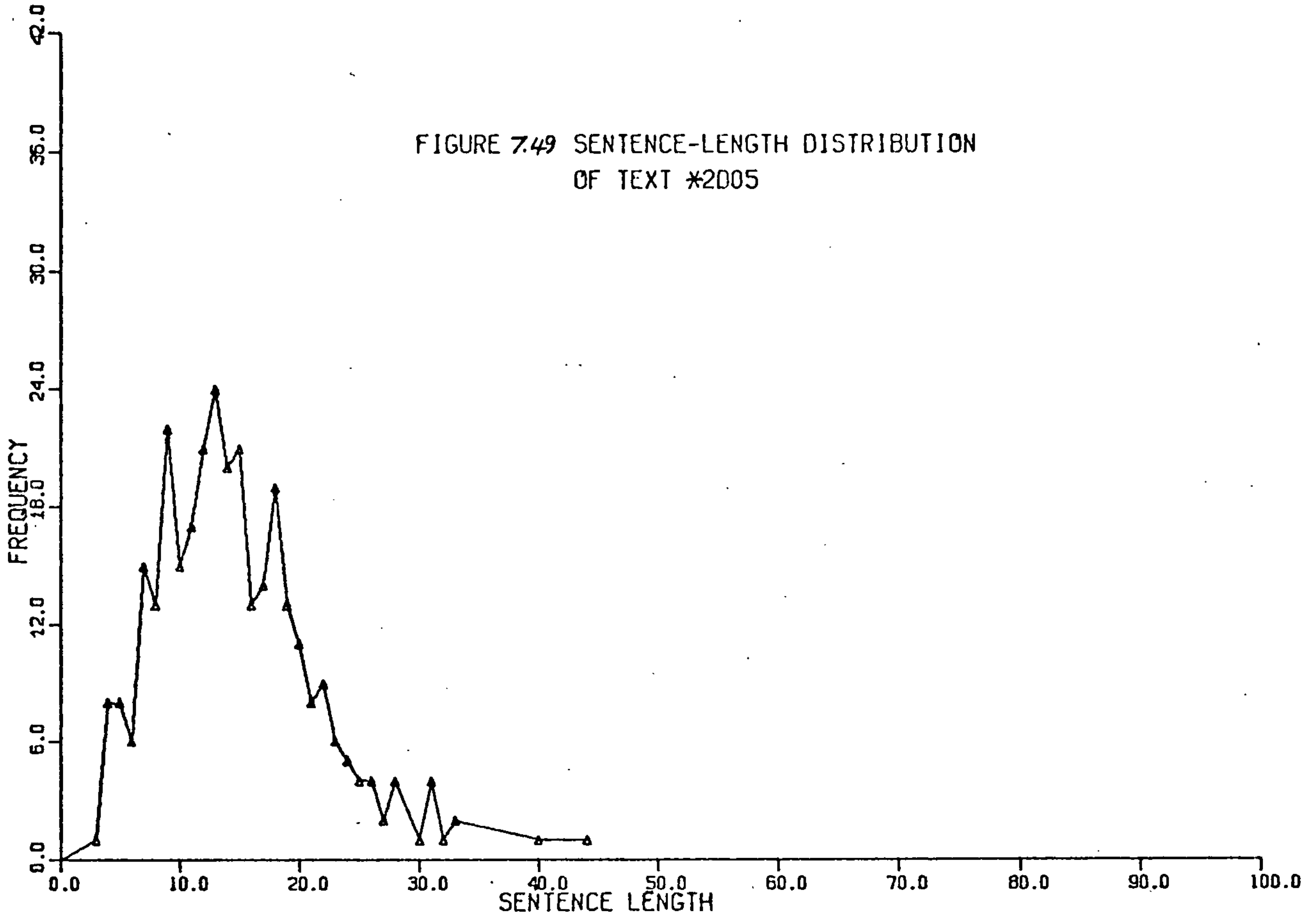


FIGURE 7.49 SENTENCE-LENGTH DISTRIBUTION
OF TEXT *2D05
FREQUENCY
SENTENCE LENGTH
0.0
6.0
12.0
18.0
24.0
30.0
36.0
42.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0

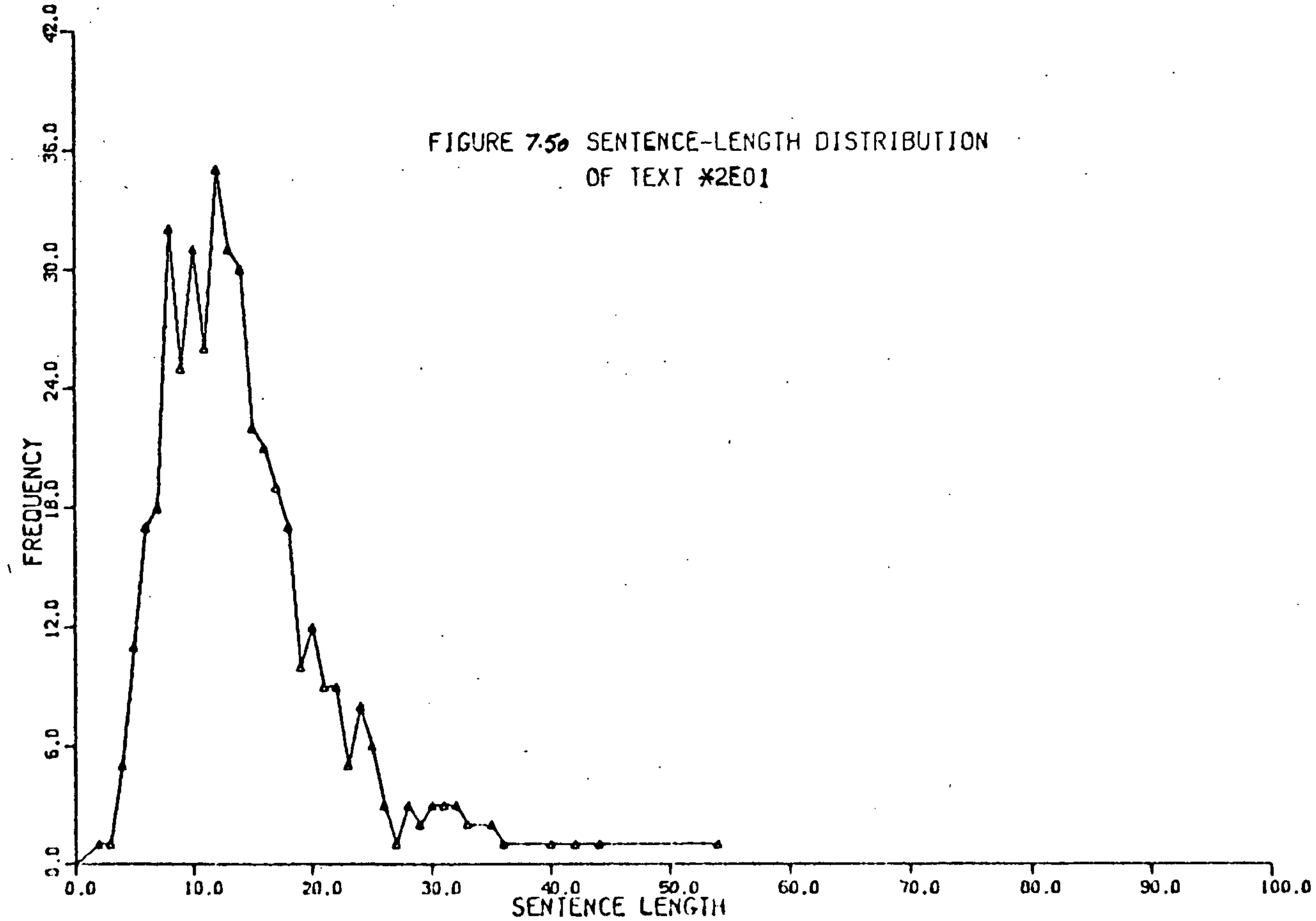


FIGURE 7.50 SENTENCE-LENGTH DISTRIBUTION
OF TEXT *2E01
FREQUENCY
SENTENCE LENGTH
0.0
6.0
12.0
18.0
24.0
30.0
36.0
42.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0

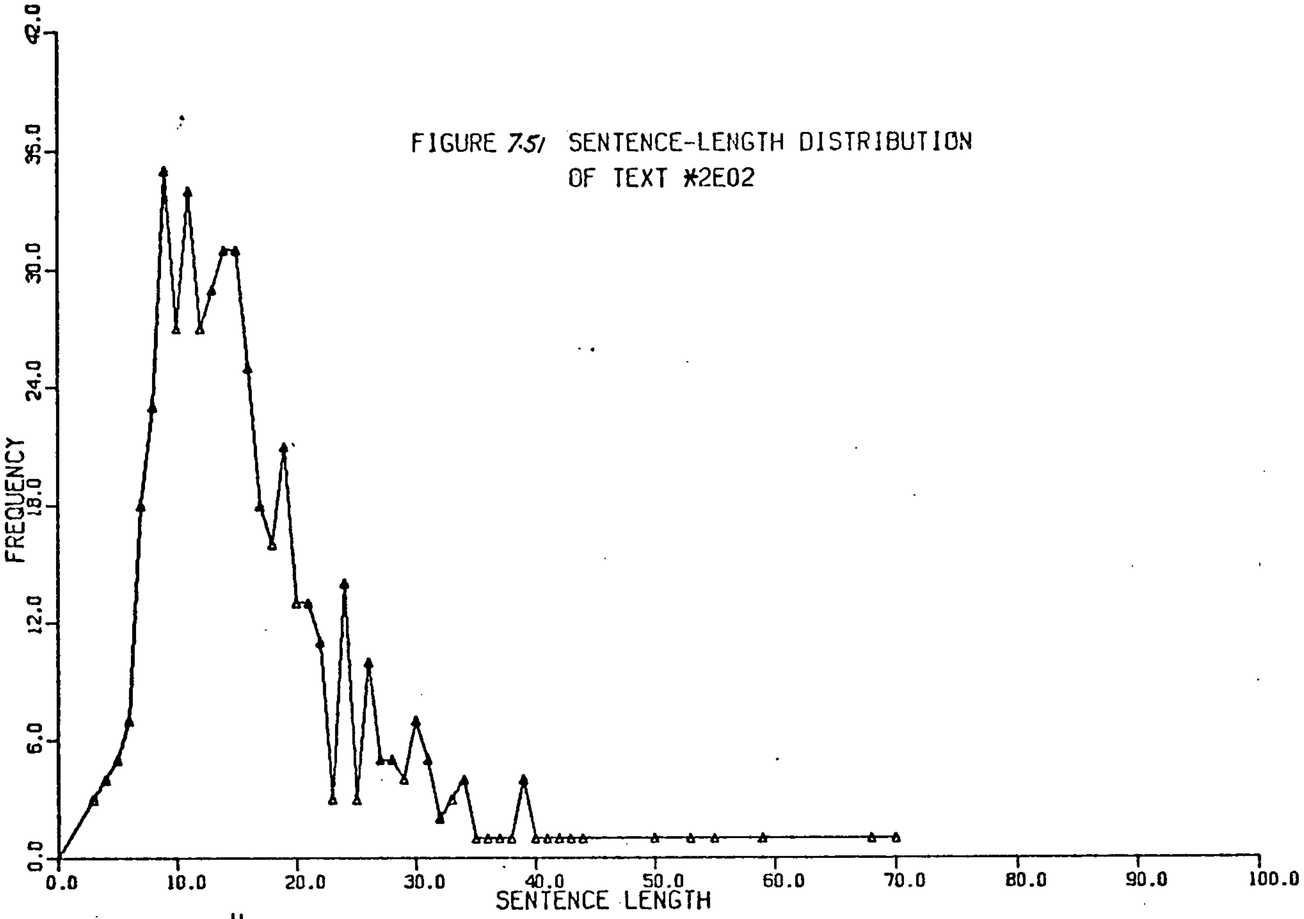


FIGURE 7.51 SENTENCE-LENGTH DISTRIBUTION
OF TEXT *2E02
FREQUENCY
SENTENCE LENGTH
0.0
6.0
12.0
18.0
24.0
30.0
36.0
42.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0

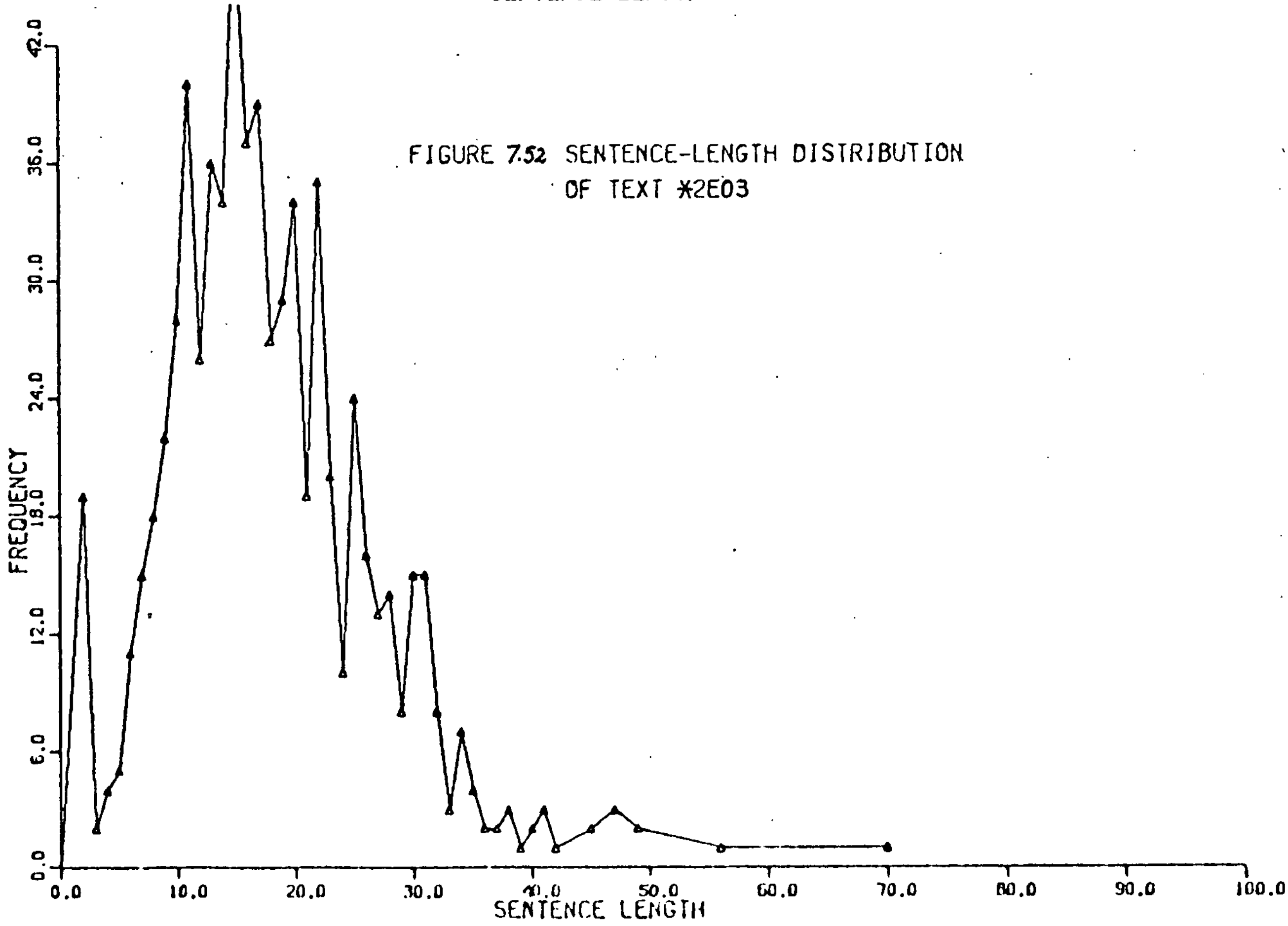


FIGURE 7.52 SENTENCE-LENGTH DISTRIBUTION
OF TEXT *2E03
FREQUENCY
SENTENCE LENGTH
0.0
6.0
12.0
18.0
24.0
30.0
36.0
42.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0

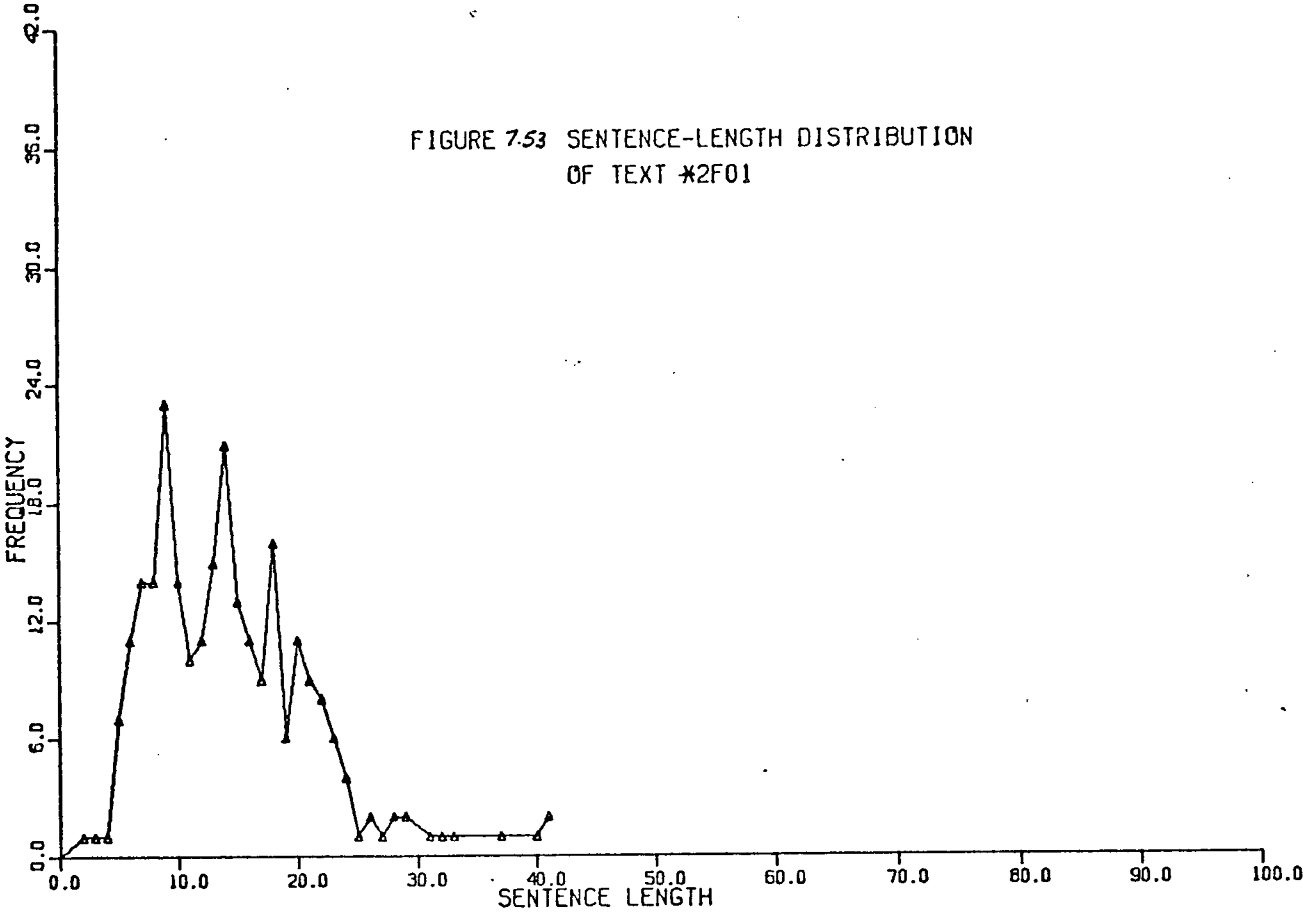


FIGURE 7.53 SENTENCE-LENGTH DISTRIBUTION OF TEXT *2F01
FREQUENCY
0.0
6.0
12.0
18.0
24.0
30.0
36.0
42.0
0.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0
SENTENCE LENGTH

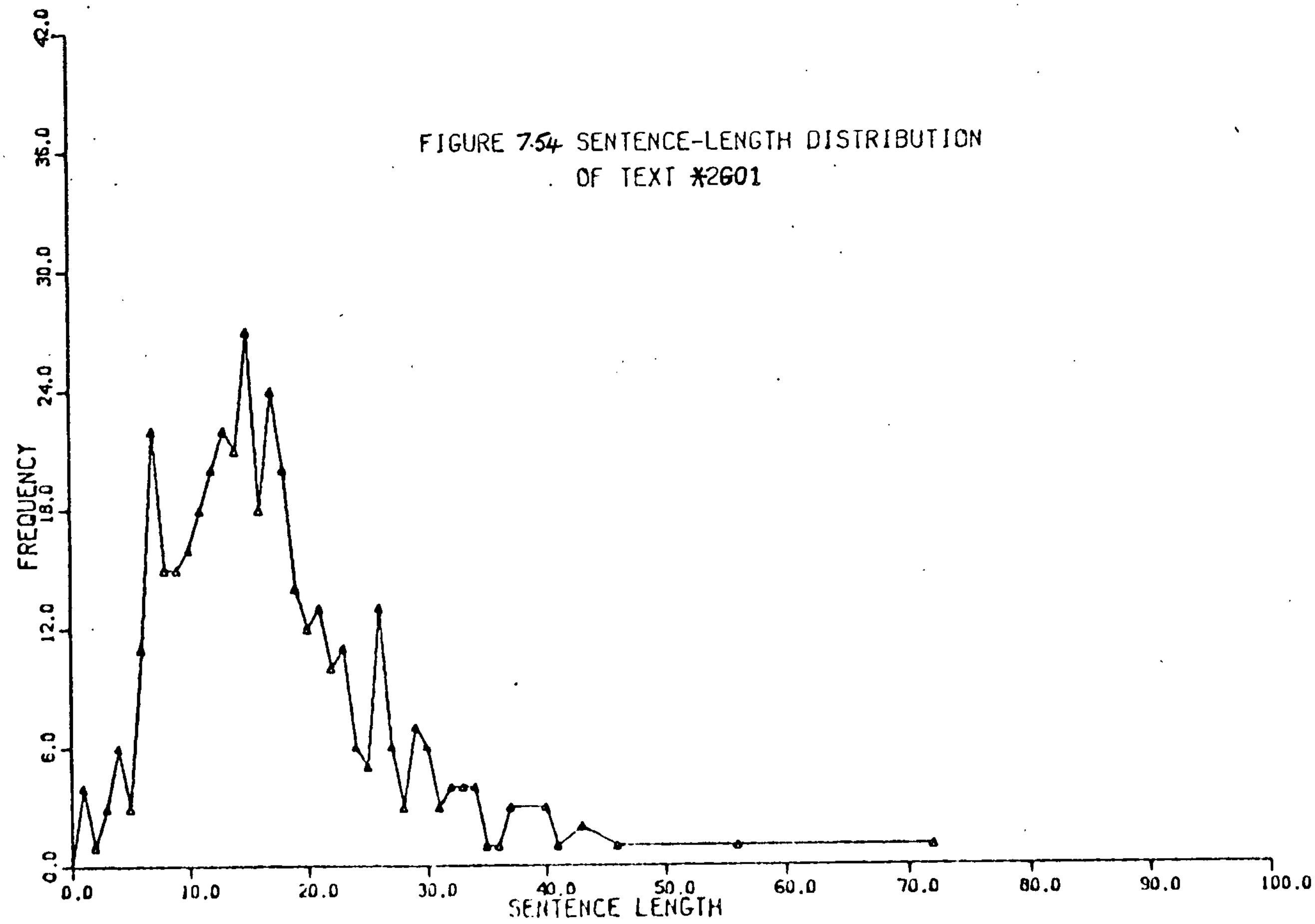


FIGURE 7.54 SENTENCE-LENGTH DISTRIBUTION OF TEXT *2G01
FREQUENCY
0.0
6.0
12.0
18.0
24.0
30.0
36.0
42.0
0.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0
SENTENCE LENGTH

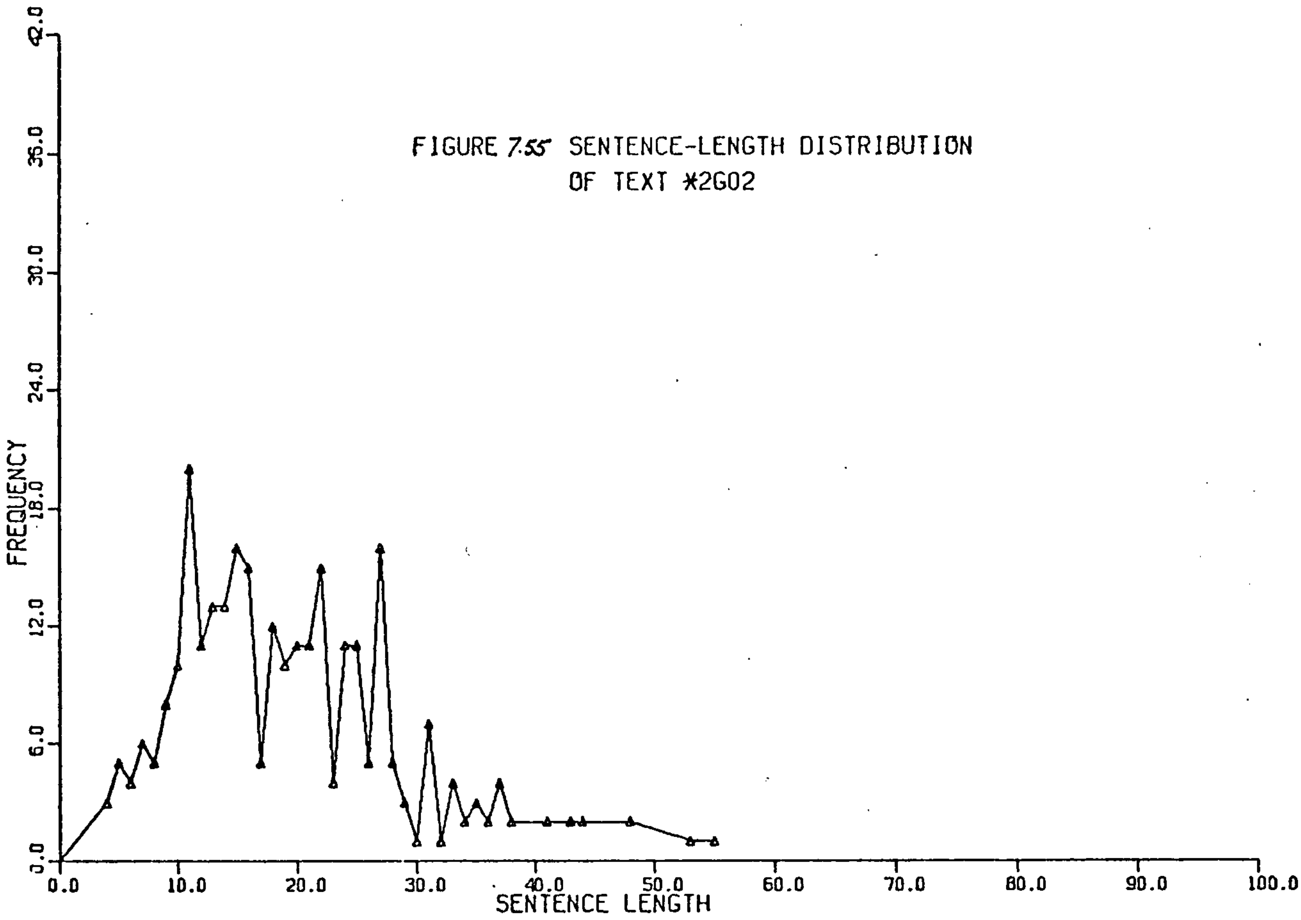


FIGURE 7.55 SENTENCE-LENGTH DISTRIBUTION
OF TEXT *2G02
FREQUENCY
0.0
6.0
12.0
18.0
24.0
30.0
36.0
42.0
0.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0
SENTENCE LENGTH

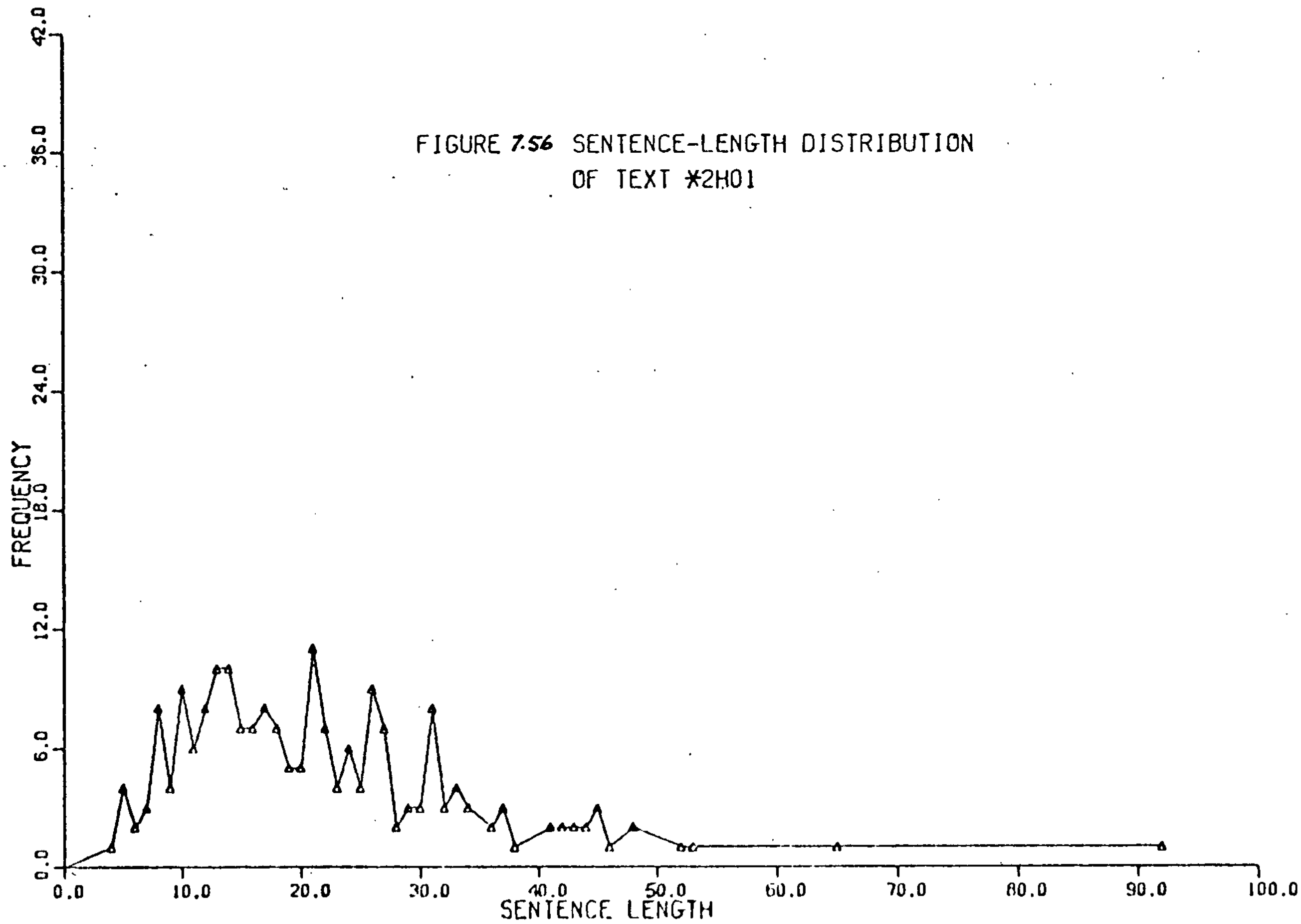


FIGURE 7.56 SENTENCE-LENGTH DISTRIBUTION
OF TEXT *2H01
FREQUENCY
0.0
6.0
12.0
18.0
24.0
30.0
36.0
42.0
0.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0
SENTENCE LENGTH

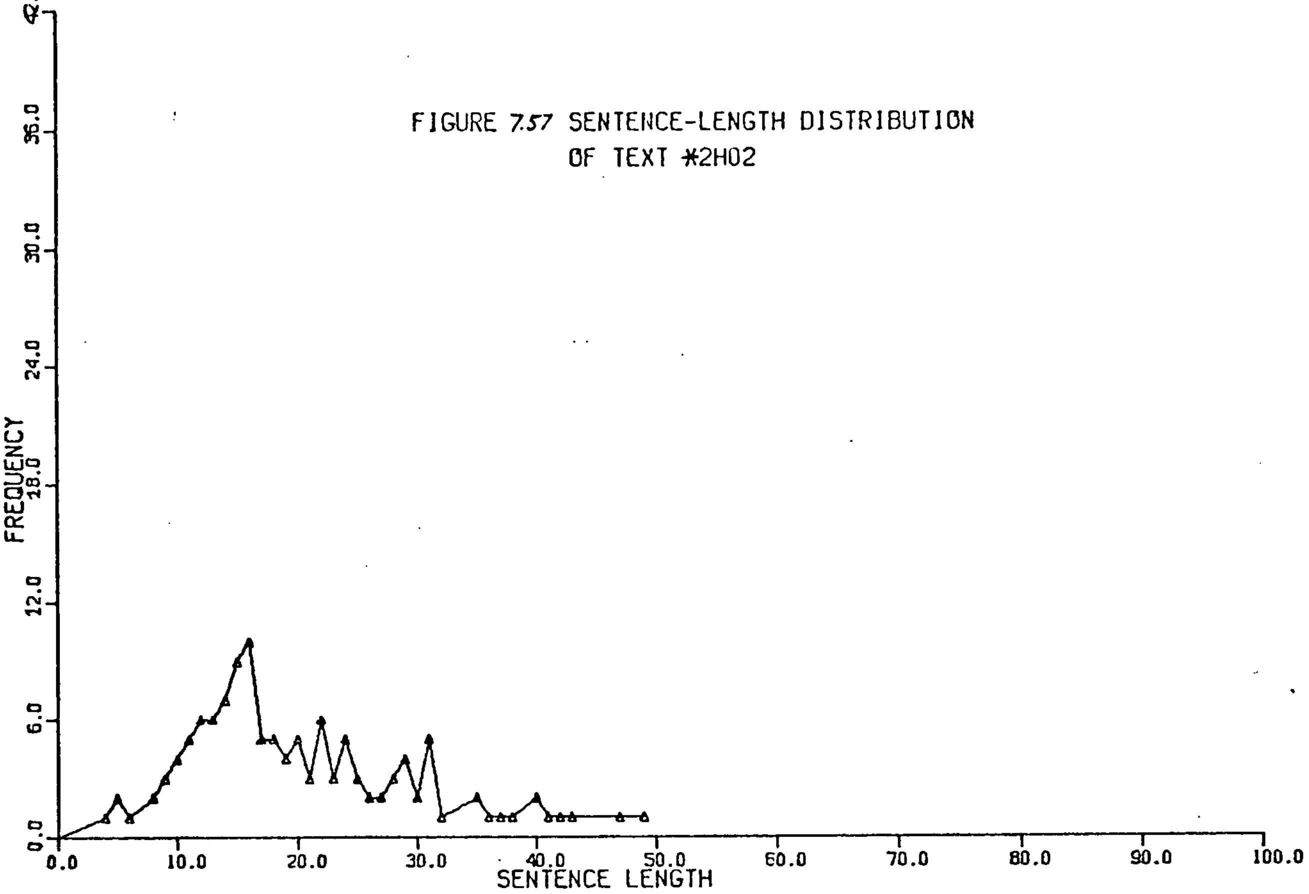


FIGURE 7.57 SENTENCE-LENGTH DISTRIBUTION
OF TEXT *2H02
FREQUENCY
SENTENCE LENGTH

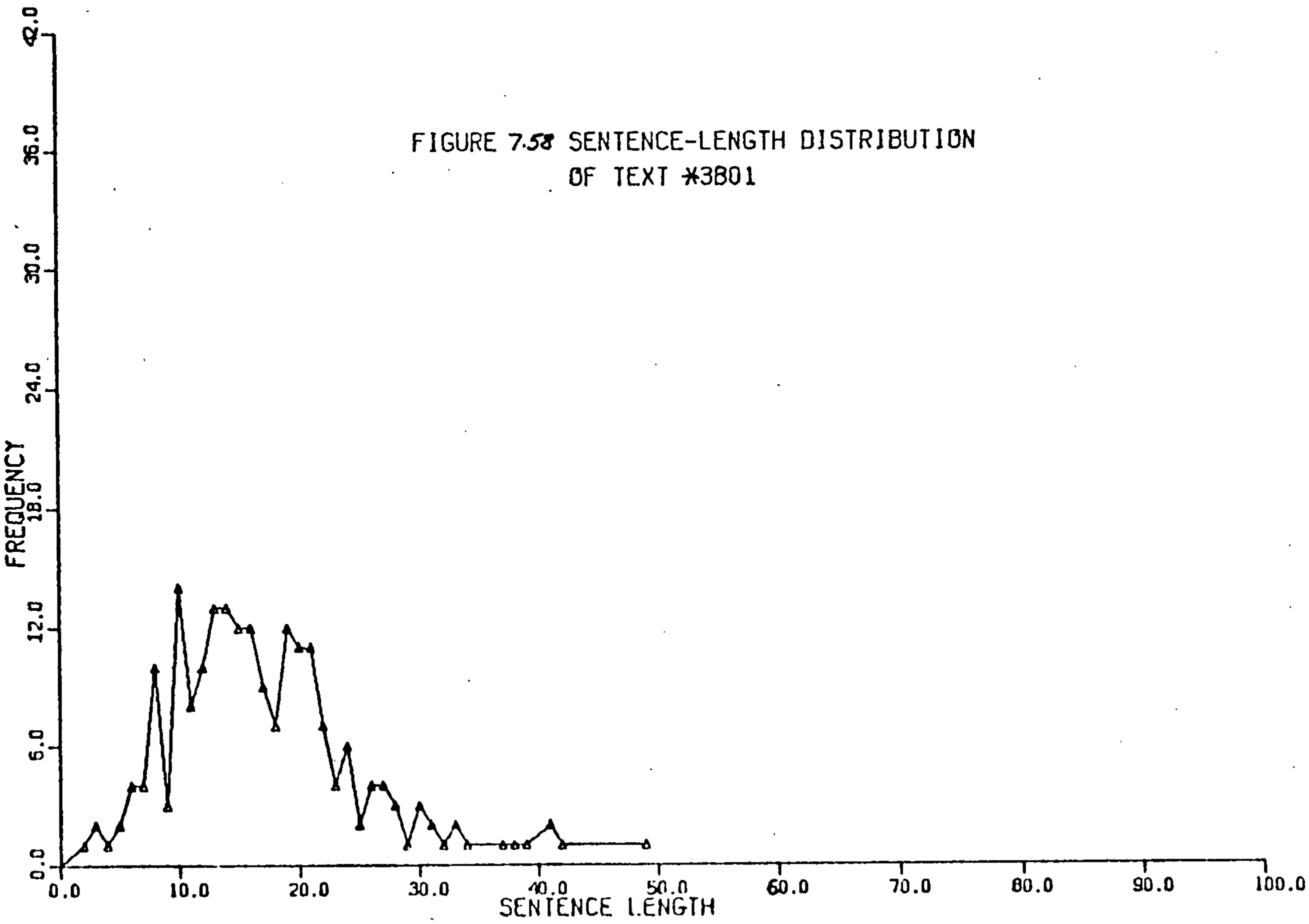


FIGURE 7.58 SENTENCE-LENGTH DISTRIBUTION
OF TEXT *3B01
FREQUENCY
SENTENCE LENGTH

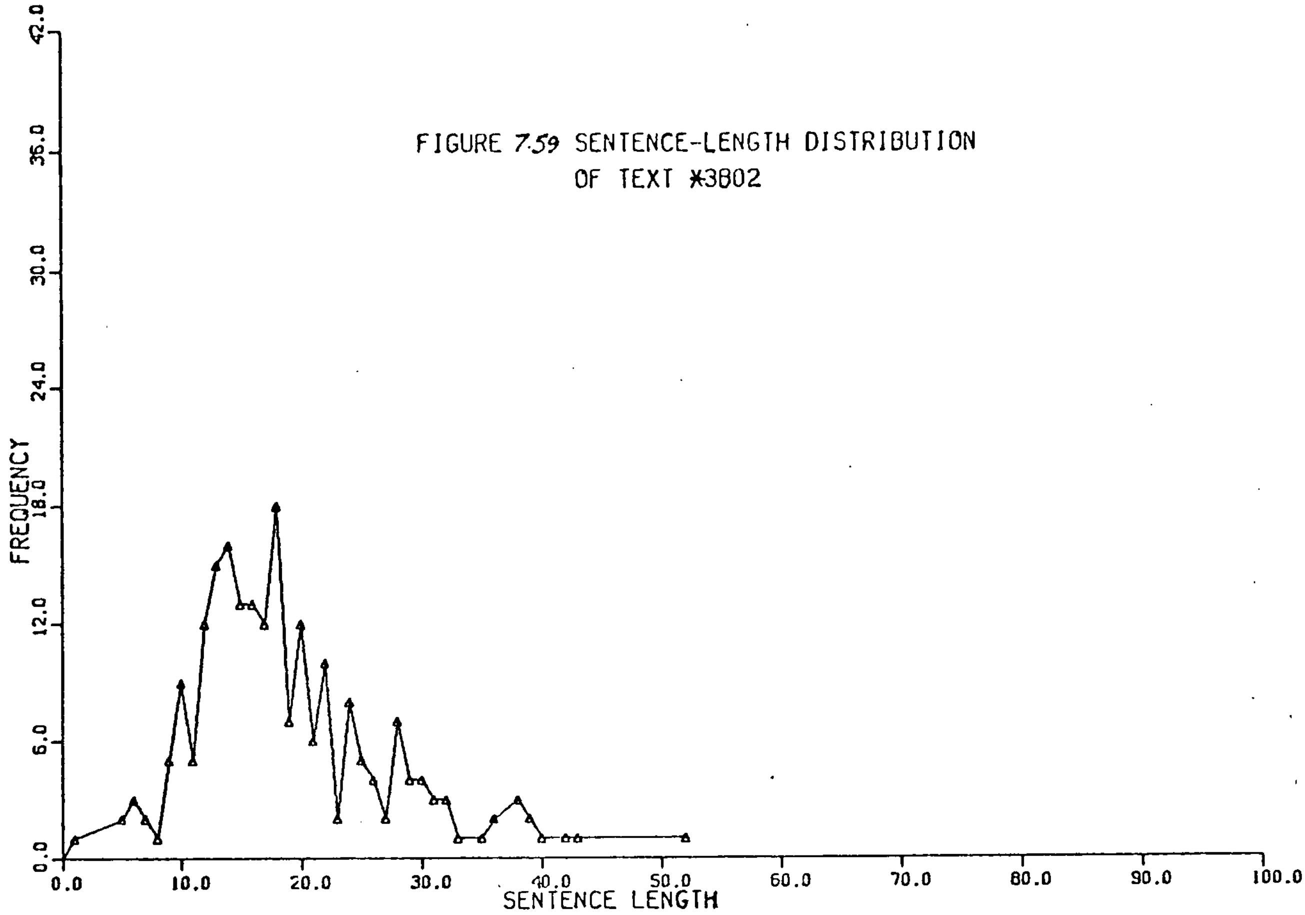


FIGURE 7.59 SENTENCE-LENGTH DISTRIBUTION
OF TEXT *3B02
FREQUENCY
SENTENCE LENGTH
0.0
6.0
12.0
18.0
24.0
30.0
36.0
42.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0

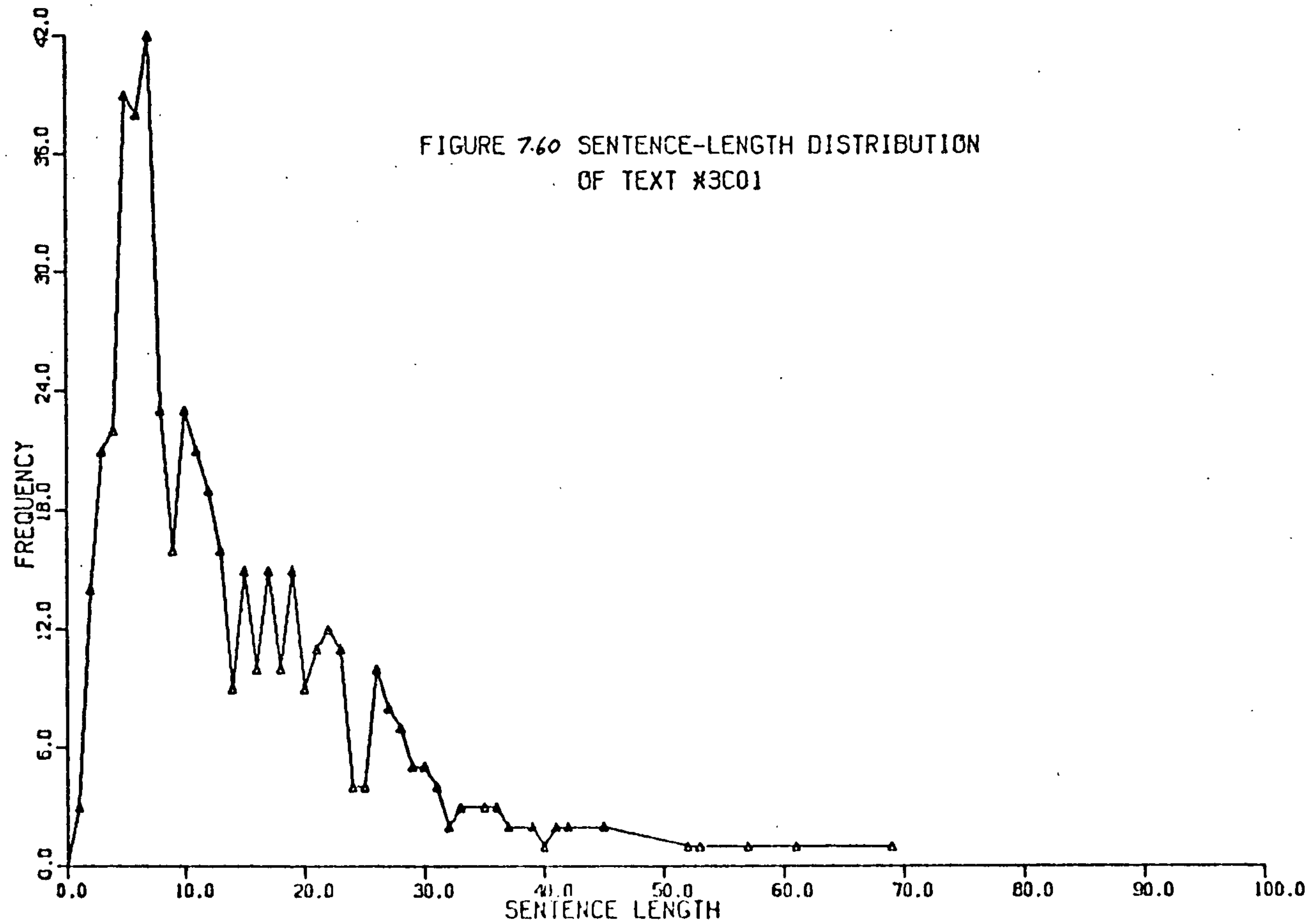


FIGURE 7.60 SENTENCE-LENGTH DISTRIBUTION
OF TEXT *3C01
FREQUENCY
SENTENCE LENGTH
0.0
6.0
12.0
18.0
24.0
30.0
36.0
42.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0

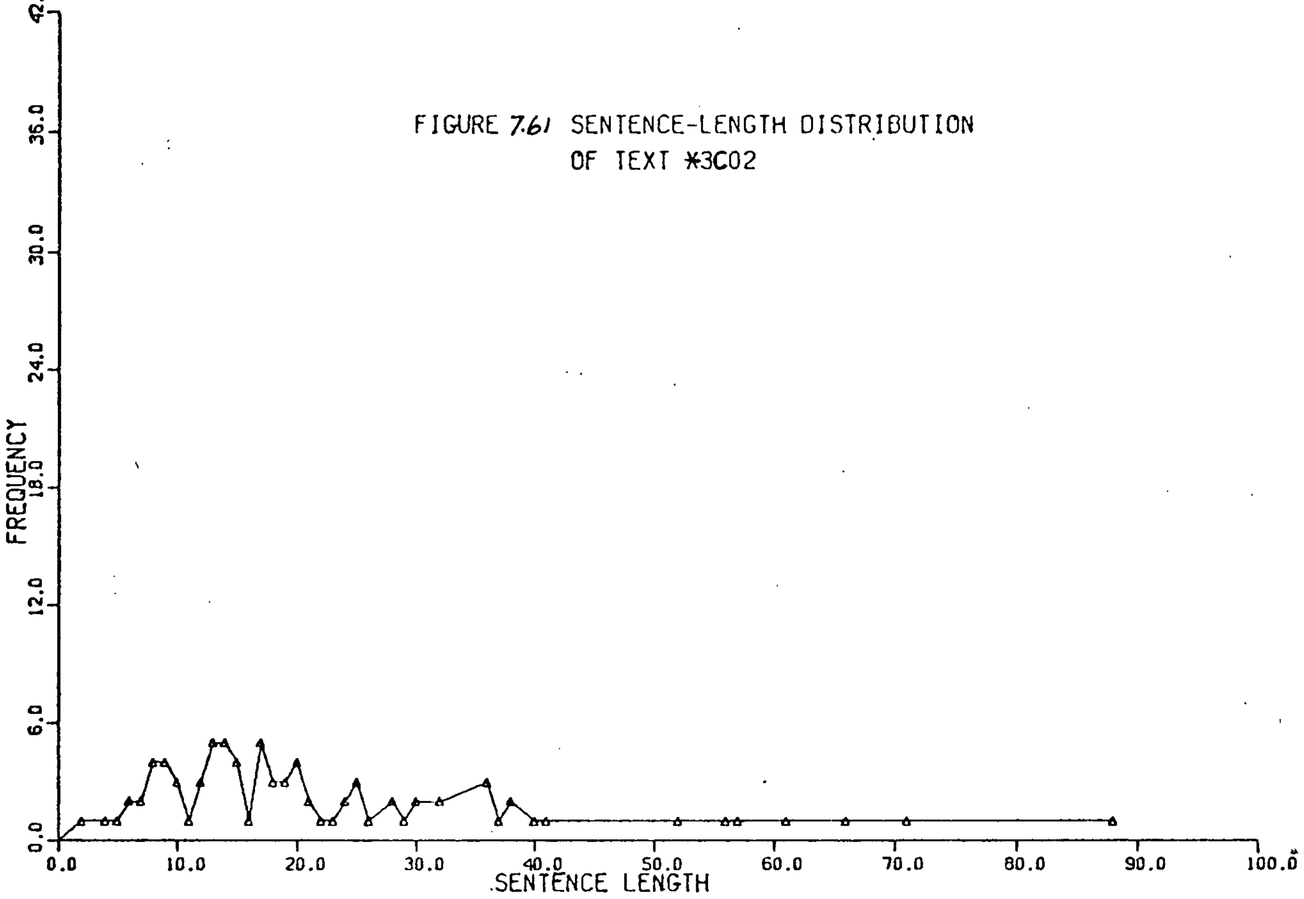


FIGURE 7.61 SENTENCE-LENGTH DISTRIBUTION
OF TEXT *3C02
FREQUENCY
0.0
6.0
12.0
18.0
24.0
30.0
36.0
42.0
0.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0
SENTENCE LENGTH

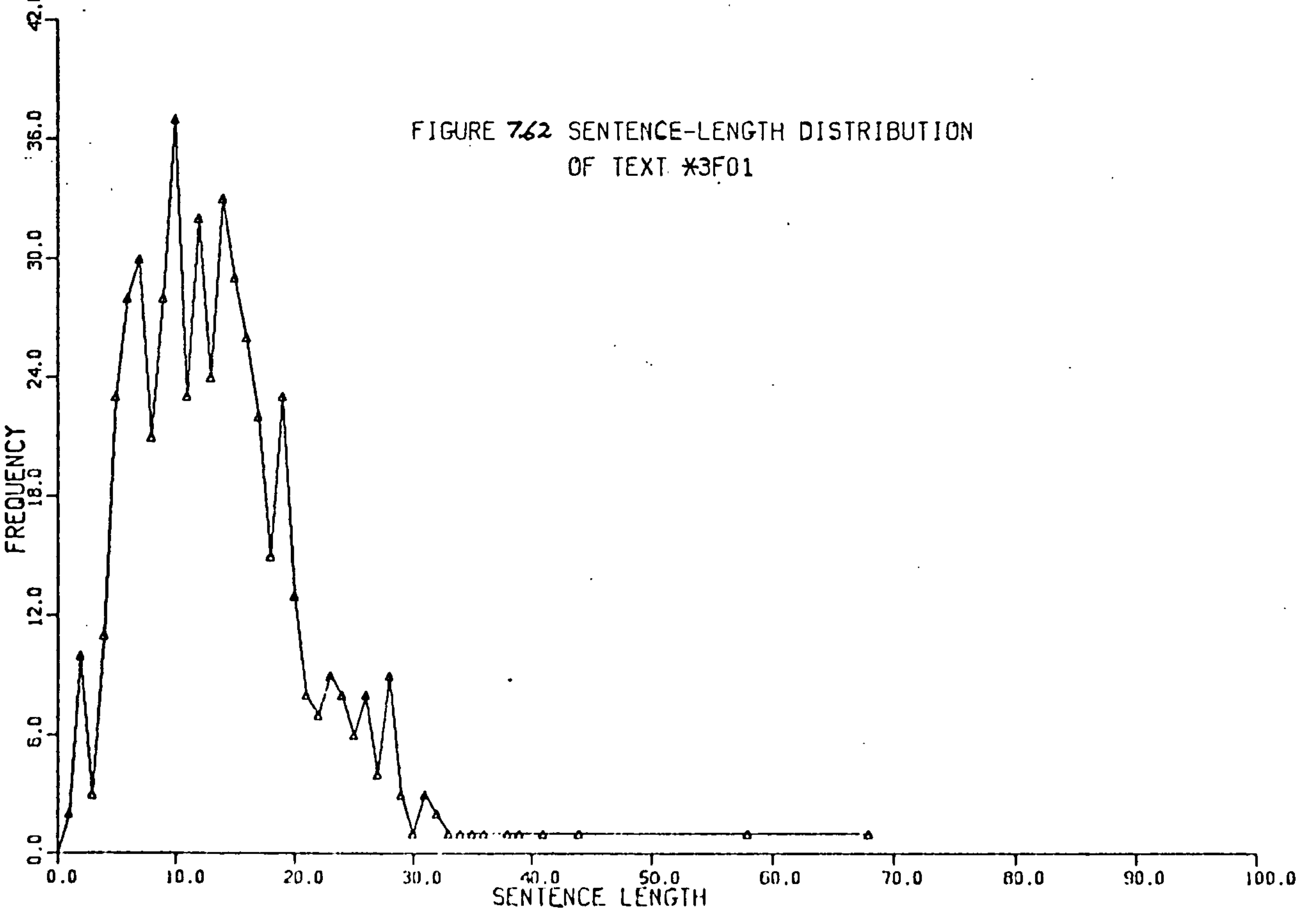


FIGURE 7.62 SENTENCE-LENGTH DISTRIBUTION
OF TEXT *3F01
FREQUENCY
0.0
6.0
12.0
18.0
24.0
30.0
36.0
42.0
0.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0
SENTENCE LENGTH

FIGURE 7.63 SENTENCE-LENGTH DISTRIBUTION OF TEXT *3G01

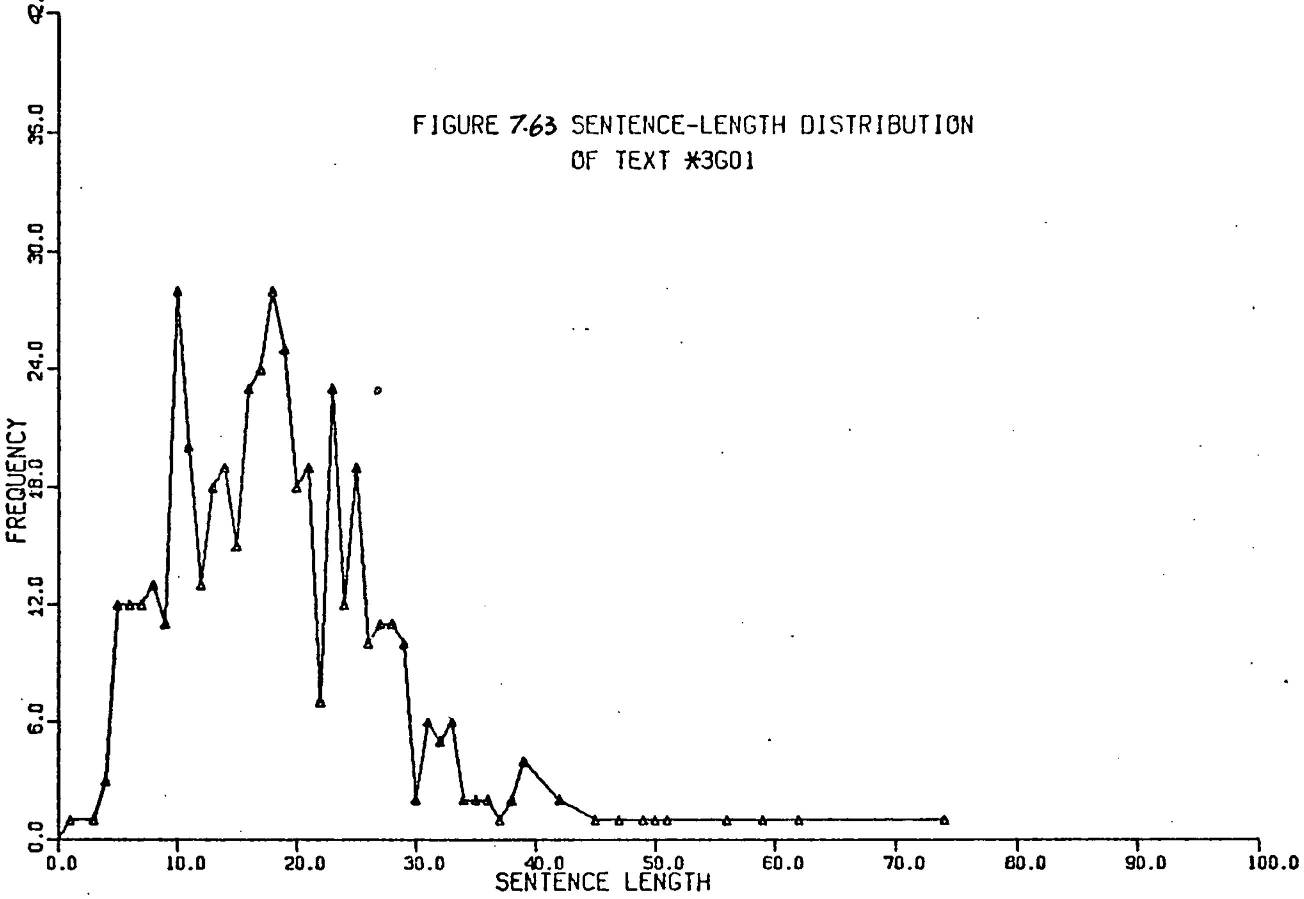
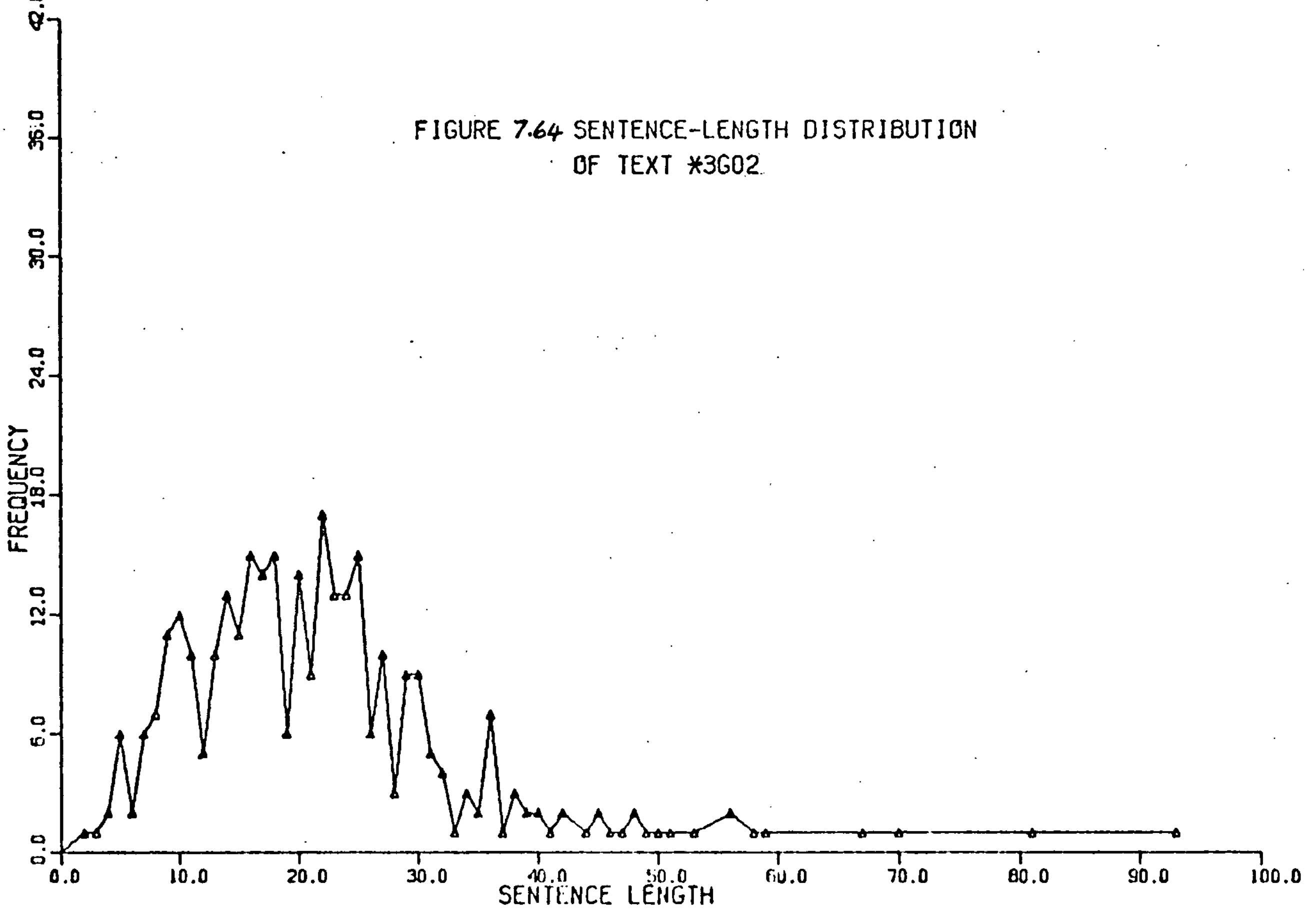


FIGURE 7.64 SENTENCE-LENGTH DISTRIBUTION OF TEXT *3G02



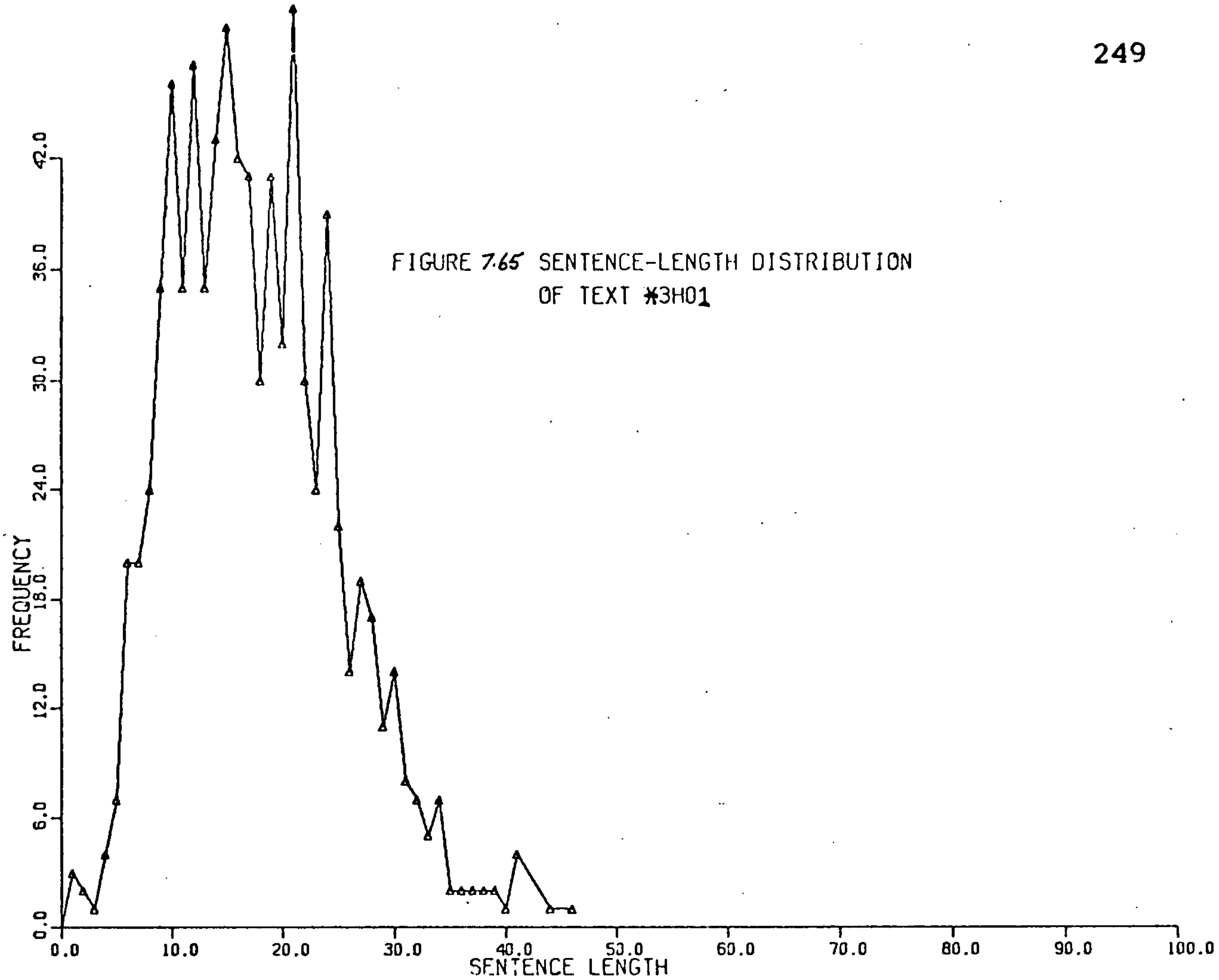


FIGURE 7.65 SENTENCE-LENGTH DISTRIBUTION OF TEXT *3H01

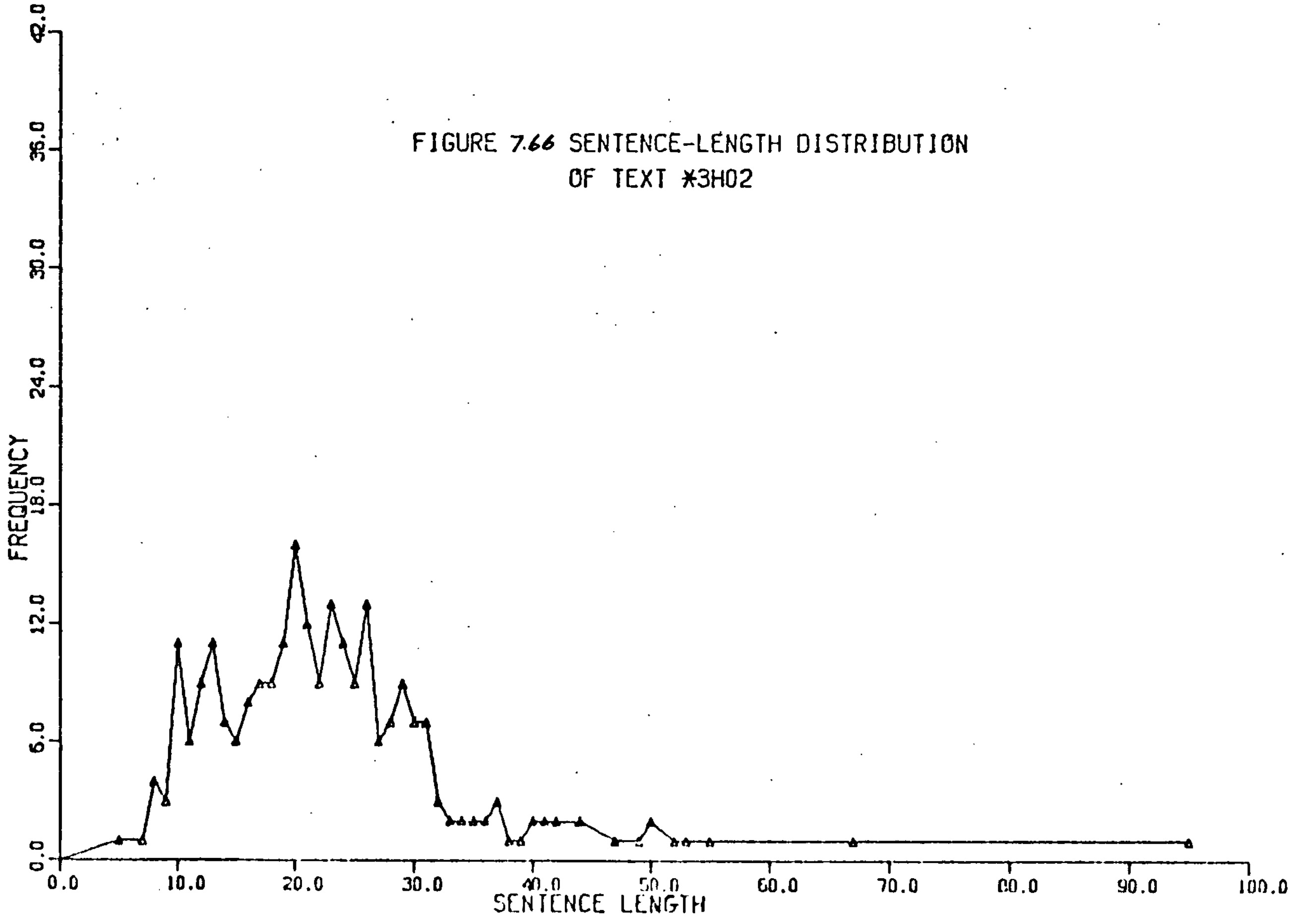


FIGURE 7.66 SENTENCE-LENGTH DISTRIBUTION OF TEXT *3H02

APPENDIX I

CHI SQUARE RESULTS OF DISTRIBUTION OF 100 MOST COMMON WORD TYPES

TABLE XXXVI   DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE GRADE LEVELS OF THE CORPUS

RANK WORD	GRADES 8	9	10	TOTAL
1.	3859.0	9071.0	4589.0	17519.
THE	3938.4	9159.7	4420.9	
	7.299	7.378	7.733	
CHI-SQUARE		8.85		
2.	1949.0	3857.0	2373.0	8177.
OF	1838.3	4275.3	2063.5	
	3.687	3.137	3.999	
CHI-SQUARE		94.03		
3.	1462.0	3539.0	1458.0	6459.
AND	1452.0	3377.0	1629.9	
	2.765	2.878	2.457	
CHI-SQUARE		25.97		
4.	1436.0	3036.0	1399.0	5921.
A	1331.1	3095.7	1494.2	
	2.716	2.510	2.357	
CHI-SQUARE		14.36		
5.	1326.0	3168.0	1427.0	5921.
TO	1331.1	3095.7	1494.2	
	2.508	2.577	2.405	
CHI-SQUARE		4.72		

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN GRADE

TABLE XXXVI   DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE GRADE LEVELS OF THE CORPUS

RANK WORD	GRADES 8	9	10	TOTAL
6.	1108.0	2515.0	1399.0	5022.
IN	1129.0	2625.7	1267.3	
	2.096	2.045	2.357	
CHI-SQUARE		18.75		
7.	891.0	2208.0	836.0	3913.
IS	879.7	2045.9	987.4	
	1.685	1.796	1.409	
CHI-SQUARE		36.22		
8.	567.0	1064.0	635.0	2266.
THAT	509.4	1184.8	571.8	
	1.073	0.865	1.070	
CHI-SQUARE		25.80		
9.	532.0	1181.0	505.0	2218.
IT	498.6	1159.7	559.7	
	1.006	0.961	0.851	
CHI-SQUARE		7.97		
10.	489.0	1278.0	398.0	2165.
ARE	486.7	1132.0	546.3	
	0.925	1.039	0.671	
CHI-SQUARE		59.13		

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN GRADE

TABLE XXXVI DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE GRADE LEVELS OF THE CORPUS

RANK / WORD	GRADES 8	9	10	TOTAL
11.	499.0	1182.0	460.0	2141.
FOR	481.3	1119.4	540.3	
	0.944	0.961	0.775	
CHI-SQUARE	16.08			
12.	545.0	1191.0	354.0	2090.
YOU	469.9	1092.7	527.4	
	1.031	0.969	0.597	
CHI-SQUARE	77.87			
13.	418.0	1102.0	351.0	1871.
BE	420.6	978.2	472.1	
	0.791	0.896	0.591	
CHI-SQUARE	46.76			
14.	405.0	899.0	485.0	1784.
AS	401.1	932.7	450.2	
	0.766	0.731	0.817	
CHI-SQUARE	3.95			
15.	369.0	1075.0	206.0	1650.
OR	370.9	862.7	416.4	
	0.698	0.874	0.347	
CHI-SQUARE	158.55			

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN GRADE

TABLE XXXVI DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE GRADE LEVELS OF THE CORPUS

RANK / WORD	GRADES 8	9	10	TOTAL
16.	298.0	870.0	317.0	1485.
WITH	333.8	776.4	374.7	
	0.564	0.708	0.534	
CHI-SQUARE	24.02			
17.	340.0	759.0	360.0	1459.
ON	328.0	762.8	368.2	
	0.643	0.617	0.607	
CHI-SQUARE	0.64			
18.	263.0	650.0	378.0	1291.
THIS	290.2	675.0	325.8	
	0.497	0.529	0.637	
CHI-SQUARE	11.85			
19.	273.0	627.0	346.0	1246.
BY	280.1	651.5	314.4	
	0.516	0.510	0.583	
CHI-SQUARE	4.27			
20.	278.0	419.0	413.0	1110.
WAS	249.5	580.4	280.1	
	0.526	0.341	0.696	
CHI-SQUARE	111.16			

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN GRADE

TABLE XXXVI DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE GRADE LEVELS OF THE CORPUS

RANK WORD	GRADES 8	9	10	TOTAL
21.	253.0	521.0	313.0	1087.
HE	244.4	568.3	274.3	
	0.479	0.424	0.527	
CHI-SQUARE		9.71		
22.	248.0	534.0	282.0	1064.
FROM	239.2	556.3	268.5	
	0.469	0.434	0.475	
CHI-SQUARE		1.90		
23.	263.0	503.0	291.0	1057.
HAVE	237.6	552.6	266.7	
	0.497	0.409	0.490	
CHI-SQUARE		9.38		
24.	241.0	526.0	287.0	1054.
AT	236.9	551.1	266.0	
	0.456	0.428	0.484	
CHI-SQUARE		2.87		
25.	257.0	484.0	251.0	992.
WHICH	223.0	518.7	250.3	
	0.486	0.394	0.423	
CHI-SQUARE		7.50		

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN GRADE

TABLE XXXVI DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE GRADE LEVELS OF THE CORPUS

RANK WORD	GRADES 8	9	10	TOTAL
26.	225.0	481.0	227.0	933.
ONE	209.7	487.8	235.4	
	0.426	0.391	0.383	
CHI-SQUARE		1.51		
27.	192.0	492.0	205.0	889.
NOT	199.9	464.8	224.3	
	0.363	0.400	0.345	
CHI-SQUARE		3.57		
28.	196.0	513.0	148.0	857.
CAN	192.7	448.1	216.3	
	0.371	0.417	0.249	
CHI-SQUARE		31.01		
29.	255.0	481.0	120.0	853.
YOUR	191.8	446.0	215.3	
	0.482	0.391	0.202	
CHI-SQUARE		65.75		
30.	230.0	452.0	171.0	833.
THEY	187.3	435.5	210.2	
	0.435	0.368	0.288	
CHI-SQUARE		17.69		

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN GRADE

TABLE XXXVI. DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE GRADE LEVELS OF THE CORPUS

RANK WORD	GRADES 8	9	10	TOTAL
31.	253.0	219.0	361.0	826.
WE	185.7	431.9	208.4	
	0.479	0.178	0.608	
CHI-SQUARE		240.98		
32.	204.0	407.0	215.0	816.
HIS	183.4	426.6	205.9	
	0.386	0.331	0.362	
CHI-SQUARE		3.61		
33.	164.0	495.0	157.0	816.
WILL	183.4	426.6	205.9	
	0.310	0.403	0.265	
CHI-SQUARE		24.64		
34.	175.0	477.0	127.0	779.
IF	175.1	407.3	196.6	
	0.331	0.388	0.214	
CHI-SQUARE		36.56		
35.	176.0	409.0	187.0	772.
AN	173.6	403.6	194.8	
	0.333	0.333	0.315	
CHI-SQUARE		0.42		

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN GRADE

TABLE XXXVI DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE GRADE LEVELS OF THE CORPUS

RANK WORD	GRADES 8	9	10	TOTAL
36.	142.0	424.0	134.0	700.
WHEN	157.4	366.0	176.6	
	0.269	0.345	0.226	
CHI-SQUARE		20.99		
37.	144.0	366.0	171.0	681.
ALL	153.1	356.1	171.8	
	0.272	0.298	0.288	
CHI-SQUARE		0.82		
38.	144.0	356.0	163.0	663.
BUT	149.0	346.6	167.3	
	0.272	0.290	0.275	
CHI-SQUARE		0.53		
39.	148.0	319.0	172.0	639.
THESE	143.7	334.1	161.3	
	0.280	0.259	0.290	
CHI-SQUARE		1.53		
40.	126.0	358.0	97.0	581.
MAY	130.6	303.8	146.6	
	0.238	0.291	0.163	
CHI-SQUARE		26.63		

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN GRADE

TABLE XXXVI. DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE GRADE LEVELS OF THE CORPUS

RANK WORD	GRADES 8	9	10	TOTAL
41.	142.0	301.0	126.0	569.
THERE	127.9	297.5	143.6	
	0.269	0.245	0.212	
CHI-SQUARE	3.75			
42.	102.0	296.0	160.0	558.
HAS	125.4	291.7	140.8	
	0.193	0.241	0.270	
CHI-SQUARE	7.06			
43.	104.0	268.0	180.0	552.
I	124.1	288.6	139.3	
	0.197	0.218	0.303	
CHI-SQUARE	16.62			
44.	139.0	269.0	134.0	542.
OTHER	121.8	283.4	136.8	
	0.263	0.219	0.226	
CHI-SQUARE	3.20			
45.	122.0	299.0	116.0	537.
SOME	120.7	280.8	135.5	
	0.231	0.243	0.195	
CHI-SQUARE	4.01			

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN GRADE

TABLE XXXVI DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE GRADE LEVELS OF THE CORPUS

RANK WORD	GRADES 8	9	10	TOTAL
46.	108.0	304.0	117.0	529.
MORE	118.9	276.6	133.5	
	0.204	0.247	0.197	
CHI-SQUARE	5.76			
47.	143.0	178.0	197.0	518.
WERE	116.5	270.8	130.7	
	0.270	0.145	0.332	
CHI-SQUARE	71.48			
48.	113.0	219.0	178.0	510.
HAD	114.7	266.6	128.7	
	0.214	0.178	0.300	
CHI-SQUARE	27.43			
49.	120.0	226.0	152.0	498.
THEIR	112.0	260.4	125.7	
	0.227	0.184	0.256	
CHI-SQUARE	10.63			
50.	117.0	310.0	64.0	491.
USED	110.4	256.7	123.9	
	0.221	0.252	0.108	
CHI-SQUARE	40.42			

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN GRADE

TABLE XXXVI DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE GRADE LEVELS OF THE CORPUS

RANK WORD	GRADES 8	9	10	TOTAL
51.	118.0	245.0	125.0	488.
MANY	109.7	255.1	123.1	
	0.223	0.199	0.211	
CHI-SQUARE		1.06		
52.	137.0	249.0	95.0	481.
SO	108.1	251.5	121.4	
	0.259	0.203	0.160	
CHI-SQUARE		13.46		
53.	122.0	253.0	101.0	476.
EACH	107.0	248.9	120.1	
	0.231	0.206	0.170	
CHI-SQUARE		5.21		
54.	117.0	225.0	123.0	465.
TWO	104.5	243.1	117.3	
	0.221	0.183	0.207	
CHI-SQUARE		3.11		
55.	91.0	265.0	107.0	463.
ABOUT	104.1	242.1	116.8	
	0.172	0.216	0.180	
CHI-SQUARE		4.64		

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN GRADE

TABLE XXXVI DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE GRADE LEVELS OF THE CORPUS

RANK WORD	GRADES 8	9	10	TOTAL
56.	109.0	266.0	64.0	439.
SHOULD	98.7	229.5	110.8	
	0.206	0.216	0.108	
CHI-SQUARE		26.63		
57.	114.0	194.0	120.0	428.
WHAT	96.2	223.8	108.0	
	0.216	0.158	0.202	
CHI-SQUARE		8.58		
58.	102.0	217.0	106.0	425.
THAN	95.5	222.2	107.2	
	0.193	0.176	0.179	
CHI-SQUARE		0.57		
59.	79.0	197.0	148.0	424.
BEEN	95.3	221.7	107.0	
	0.149	0.160	0.249	
CHI-SQUARE		21.26		
60.	96.0	237.0	90.0	423.
INTO	95.1	221.2	106.7	
	0.182	0.193	0.152	
CHI-SQUARE		3.77		

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN GRADE

TABLE XXXVI DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE GRADE LEVELS OF THE CORPUS

RANK WORD	GRADES 8	9	10	TOTAL
61.	109.0	220.0	92.0	421.
THEM	94.6	220.1	106.2	
	0.206	0.179	0.155	
CHI-SQUARE		4.09		
62.	97.0	253.0	67.0	417.
USE	93.7	218.0	105.2	
	0.183	0.206	0.113	
CHI-SQUARE		19.61		
63.	107.0	255.0	49.0	411.
MAKE	92.4	214.9	103.7	
	0.202	0.207	0.083	
CHI-SQUARE		38.66		
64.	93.0	234.0	79.0	406.
DO	91.3	212.3	102.5	
	0.176	0.190	0.133	
CHI-SQUARE		7.63		
65.	81.0	240.0	82.0	403.
UP	90.6	210.7	101.7	
	0.153	0.195	0.138	
CHI-SQUARE		8.90		

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN GRADE

TABLE XXXVI DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE GRADE LEVELS OF THE CORPUS

RANK WORD	GRADES 8	9	10	TOTAL
66.	77.0	214.0	109.0	400.
SUCH	89.9	209.1	100.9	
	0.146	0.174	0.184	
CHI-SQUARE		2.61		
67.	64.0	238.0	98.0	400.
THEN	89.9	209.1	100.9	
	0.121	0.194	0.165	
CHI-SQUARE		11.54		
68.	70.0	218.0	105.0	393.
TIME	88.4	205.5	99.2	
	0.132	0.177	0.177	
CHI-SQUARE		4.92		
69.	76.0	185.0	125.0	386.
ITS	86.8	201.8	97.4	
	0.144	0.150	0.211	
CHI-SQUARE		10.56		
70.	94.0	162.0	90.0	369.
WOULD	83.0	192.9	93.1	
	0.178	0.132	0.152	
CHI-SQUARE		6.53		

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN GRADE

TABLE XXXVI DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE GRADE LEVELS OF THE CORPUS

RANK WORD	GRADES 8	9	10	TOTAL
71.	85.0	204.0	79.0	368.
HOW	82.7	192.4	92.9	
	0.161	0.166	0.133	
CHI-SQUARE		2.83		
72.	118.0	98.0	152.0	366.
NUMBER	82.3	191.4	92.4	
	0.223	0.080	0.256	
CHI-SQUARE		99.57		
73.	96.0	210.0	57.0	363.
MADE	81.6	189.8	91.6	
	0.182	0.171	0.096	
CHI-SQUARE		17.76		
74.	85.0	201.0	72.0	358.
OUT	80.5	187.2	90.3	
	0.161	0.163	0.121	
CHI-SQUARE		5.00		
75.	76.0	173.0	104.0	353.
MOST	79.4	184.6	89.1	
	0.144	0.141	0.175	
CHI-SQUARE		3.37		

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN GRADE

TABLE XXXVI DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE GRADE LEVELS OF THE CORPUS

RANK WORD	GRADES 8	9	10	TOTAL
76.	102.0	161.0	88.0	351.
ONLY	78.9	183.5	88.6	
	0.193	0.131	0.148	
CHI-SQUARE		9.52		
77.	70.0	192.0	89.0	349.
NO	78.5	182.5	88.1	
	0.132	0.156	0.150	
CHI-SQUARE		1.42		
78.	73.0	216.0	53.0	342.
MUST	76.9	178.8	86.3	
	0.138	0.176	0.089	
CHI-SQUARE		20.78		
79.	52.0	198.0	90.0	340.
WATER	76.4	177.8	85.8	
	0.098	0.161	0.152	
CHI-SQUARE		10.32		
80.	60.0	174.0	78.0	312.
ALSO	70.1	163.1	78.7	
	0.113	0.142	0.131	
CHI-SQUARE		2.20		

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN GRADE

TABLE XXXVI DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE GRADE LEVELS OF THE CORPUS

RANK WORD	GRADES 8	9	10	TOTAL
81.	73.0	142.0	93.0	308.
FIRST	69.2	161.0	77.7	
	0.138	0.115	0.157	
CHI-SQUARE		5.46		
82.	73.0	157.0	76.0	306.
VERY	68.8	160.0	77.2	
	0.138	0.128	0.128	
CHI-SQUARE		0.33		
83.	93.0	168.0	42.0	303.
GOOD	68.1	158.4	76.5	
	0.176	0.137	0.071	
CHI-SQUARE		25.20		
84.	64.0	160.0	71.0	295.
HIM	65.3	154.2	74.4	
	0.121	0.130	0.120	
CHI-SQUARE		0.46		
85.	70.0	137.0	79.0	286.
SAME	64.3	149.5	72.2	
	0.132	0.111	0.133	
CHI-SQUARE		2.20		

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN GRADE

TABLE XXXVI DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE GRADE LEVELS OF THE CORPUS

RANK WORD	GRADES 8	9	10	TOTAL
86.	69.0	95.0	96.0	260.
COULD	58.5	135.9	65.6	
	0.131	0.077	0.162	
CHI-SQUARE		28.31		
87.	64.0	132.0	64.0	260.
WHO	58.5	135.9	65.6	
	0.121	0.107	0.108	
CHI-SQUARE		0.68		
88.	58.0	126.0	71.0	255.
ANY	57.3	133.3	64.3	
	0.110	0.102	0.120	
CHI-SQUARE		1.10		
89.	60.0	144.0	50.0	254.
BECAUSE	57.1	132.8	64.1	
	0.113	0.117	0.084	
CHI-SQUARE		4.19		
90.	67.0	133.0	53.0	253.
SEE	56.9	132.3	63.8	
	0.127	0.108	0.089	
CHI-SQUARE		3.65		

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN GRADE

TABLE XXXVI DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE GRADE LEVELS OF THE CORPUS

RANK WORD	GRADES 8	9	10	TOTAL
91.	64.0	129.0	52.0	245.
LIKE	55.1	128.1	61.8	
	0.121	0.105	0.088	
CHI-SQUARE		3.01		
92.	55.0	116.0	68.0	239.
MUCH	53.7	125.0	60.3	
	0.104	0.094	0.115	
CHI-SQUARE		1.65		
93.	60.0	104.0	72.0	236.
PEOPLE	53.1	123.4	59.6	
	0.113	0.085	0.121	
CHI-SQUARE		6.56		
94.	53.0	111.0	70.0	234.
CALLED	52.6	122.3	59.0	
	0.100	0.090	0.118	
CHI-SQUARE		3.09		
95.	74.0	123.0	36.0	233.
PLACE	52.4	121.8	58.8	
	0.140	0.100	0.061	
CHI-SQUARE		17.77		

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN GRADE

TABLE XXXVI DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE GRADE LEVELS OF THE CORPUS

RANK WORD	GRADES 8	9	10	TOTAL
96.	44.0	136.0	52.0	232.
THROUGH	52.2	121.3	58.5	
	0.083	0.111	0.088	
CHI-SQUARE		3.79		
97.	43.0	163.0	26.0	232.
WORK	52.2	121.3	58.5	
	0.081	0.133	0.044	
CHI-SQUARE		34.03		
98.	38.0	85.0	105.0	228.
NEW	51.3	119.2	57.5	
	0.072	0.069	0.177	
CHI-SQUARE		52.40		
99.	48.0	117.0	58.0	223.
SMALL	50.1	116.6	56.3	
	0.091	0.095	0.098	
CHI-SQUARE		0.15		
100.	52.0	123.0	45.0	220.
OVER	49.5	115.0	55.5	
	0.098	0.100	0.076	
CHI-SQUARE		2.68		

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN GRADE

TABLE XXXVII — DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF THE CORPUS

RANK / WORD	B	C	D	E	F	G	H	TOTAL
1.	1463.0	2799.0	2652.0	2787.0	1263.0	3266.0	3289.0	17519.
THE	1501.8	3003.0	3670.4	2332.3	1327.0	2815.7	2876.9	
	7.259	6.945	5.384	8.904	7.092	8.643	8.519	
CHI-SQUARE	520.19							
2.	607.0	1189.0	1367.0	845.0	757.0	1636.0	1782.0	8177.
OF	701.0	1401.6	1713.2	1088.6	619.4	1314.2	1342.8	
	3.012	2.950	2.775	2.700	4.251	4.330	4.616	
CHI-SQUARE	422.35							
3.	471.0	1257.0	1653.0	855.0	287.0	746.0	1190.0	6459.
AND	553.7	1107.1	1353.2	859.9	489.2	1038.1	1060.7	
	2.337	3.119	3.356	2.732	1.612	1.974	3.082	
CHI-SQUARE	280.64							
4.	502.0	992.0	1369.0	797.0	525.0	1033.0	703.0	5921.
A	507.6	1014.9	1240.5	788.3	448.5	951.6	972.3	
	2.491	2.462	2.779	2.546	2.948	2.734	1.821	
CHI-SQUARE	108.59							
5.	599.0	975.0	1494.0	695.0	421.0	859.0	878.0	5921.
TO	507.6	1014.9	1240.5	788.3	448.5	951.6	972.3	
	2.972	2.419	3.033	2.220	2.364	2.273	2.274	
CHI-SQUARE	100.72							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXVII — DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF THE CORPUS

RANK / WORD	B	C	D	E	F	G	H	TOTAL
6.	411.0	777.0	1046.0	620.0	358.0	799.0	1011.0	5022.
IN	430.5	860.8	1052.2	668.6	380.4	807.1	824.7	
	2.039	1.928	2.124	1.981	2.010	2.114	2.619	
CHI-SQUARE	56.11							
7.	350.0	406.0	903.0	797.0	427.0	595.0	441.0	3913.
IS	335.4	670.7	819.8	520.9	296.4	628.9	642.6	
	1.737	1.007	1.833	2.546	2.398	1.575	1.142	
CHI-SQUARE	382.47							
8.	200.0	482.0	414.0	204.0	307.0	425.0	234.0	2266.
THAT	194.3	388.4	474.7	301.7	171.6	364.2	372.1	
	0.992	1.196	0.840	0.652	1.724	1.125	0.606	
CHI-SQUARE	230.28							
9.	176.0	478.0	503.0	337.0	140.0	352.0	232.0	2218.
IT	190.1	380.2	464.7	295.3	168.0	356.5	364.2	
	0.873	1.186	1.021	1.077	0.786	0.932	0.601	
CHI-SQUARE	87.99							
10.	213.0	196.0	700.0	364.0	181.0	288.0	241.0	2165.
ARE	185.6	371.1	453.6	288.2	164.0	348.0	355.5	
	1.057	0.486	1.421	1.163	1.016	0.762	0.624	
CHI-SQUARE	289.44							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXVII. DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF THE CORPUS

RANK WORD	B	C	D	E	F	G	H	TOTAL
11.	258.0	317.0	625.0	265.0	145.0	205.0	326.0	2141.
FOR	183.5	367.0	448.6	285.0	162.2	344.1	351.6	
	1.280	0.787	1.269	0.847	0.814	0.543	0.844	
CHI-SQUARE	167.74							
12.	366.0	365.0	633.0	140.0	187.0	346.0	53.0	2090.
YOU	179.2	358.2	437.9	278.2	158.3	335.9	343.2	
	1.816	0.906	1.285	0.447	1.050	0.916	0.137	
CHI-SQUARE	601.48							
13.	191.0	185.0	663.0	325.0	121.0	234.0	152.0	1871.
BE	160.4	320.7	392.0	249.1	141.7	300.7	307.2	
	0.948	0.459	1.346	1.038	0.679	0.619	0.394	
CHI-SQUARE	370.04							
14.	135.0	309.0	406.0	214.0	148.0	274.0	303.0	1789.
AS	153.4	306.7	374.8	238.2	135.5	287.5	293.8	
	0.670	0.767	0.824	0.684	0.831	0.725	0.785	
CHI-SQUARE	9.34							
15.	138.0	166.0	636.0	319.0	71.0	198.0	122.0	1650.
OR	141.4	282.8	345.7	219.7	125.0	265.2	271.0	
	0.685	0.412	1.291	1.019	0.399	0.524	0.316	
CHI-SQUARE	459.29							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXVII DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF THE CORPUS

RANK WORD	B	C	D	E	F	G	H	TOTAL
16.	98.0	284.0	364.0	231.0	84.0	224.0	200.0	1485.
WITH	127.3	254.5	311.1	197.7	112.5	238.7	243.9	
	0.486	0.705	0.739	0.738	0.472	0.593	0.518	
CHI-SQUARE	40.75							
17.	165.0	210.0	276.0	215.0	97.0	244.0	252.0	1459.
ON	125.1	250.1	305.7	194.2	110.5	234.5	239.6	
	0.819	0.521	0.560	0.687	0.545	0.646	0.653	
CHI-SQUARE	26.95							
18.	163.0	136.0	148.0	211.0	165.0	230.0	238.0	1291.
THIS	110.7	221.3	270.5	171.9	97.8	207.5	212.0	
	0.809	0.337	0.300	0.674	0.927	0.609	0.616	
CHI-SQUARE	173.81							
19.	126.0	166.0	199.0	167.0	112.0	198.0	278.0	1246.
BY	106.8	213.6	261.0	165.9	94.4	200.3	204.6	
	0.625	0.412	0.404	0.534	0.629	0.524	0.720	
CHI-SQUARE	58.44							
20.	24.0	450.0	39.0	41.0	59.0	136.0	361.0	1110.
WAS	95.2	190.3	232.6	147.8	84.1	178.4	182.3	
	0.119	1.117	0.079	0.131	0.331	0.360	0.935	
CHI-SQUARE	838.81							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXVII DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF THE CORPUS

RANK / WORD	B	C	D	E	F	G	H	TOTAL
21.	79.0	520.0	187.0	14.0	42.0	137.0	108.0	1087.
HE	93.2	186.3	227.7	144.7	82.3	174.7	178.5	
	0.392	1.290	0.380	0.045	0.236	0.363	0.280	
CHI-SQUARE	780.81							
22.	90.0	148.0	189.0	137.0	55.0	242.0	203.0	1064.
FROM	91.2	182.4	222.9	141.7	80.6	171.0	174.7	
	0.447	0.367	0.384	0.438	0.309	0.640	0.526	
CHI-SQUARE	53.98							
23.	145.0	180.0	290.0	76.0	114.0	138.0	114.0	1057.
HAVE	90.6	181.2	221.5	140.7	80.1	169.9	173.6	
	0.719	0.447	0.589	0.243	0.640	0.365	0.295	
CHI-SQUARE	124.45							
24.	112.0	224.0	207.0	100.0	49.0	222.0	140.0	1054.
AT	90.4	180.7	220.8	140.3	79.8	169.4	173.1	
	0.556	0.556	0.420	0.319	0.275	0.588	0.363	
CHI-SQUARE	62.59							
25.	70.0	106.0	198.0	129.0	84.0	200.0	205.0	992.
WHICH	85.0	170.0	207.8	132.1	75.1	159.4	162.9	
	0.347	0.263	0.402	0.412	0.472	0.529	0.531	
CHI-SQUARE	49.56							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXVII DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF THE CORPUS

RANK / WORD	B	C	D	E	F	G	H	TOTAL
26.	68.0	141.0	182.0	108.0	121.0	180.0	133.0	933.
ONE	80.0	159.9	195.5	124.2	70.7	150.0	153.2	
	0.337	0.350	0.369	0.345	0.679	0.476	0.344	
CHI-SQUARE	51.61							
27.	63.0	201.0	236.0	77.0	88.0	122.0	102.0	889.
NOT	76.2	152.4	186.3	118.4	67.3	142.9	146.0	
	0.313	0.499	0.479	0.246	0.494	0.323	0.264	
CHI-SQUARE	68.18							
28.	84.0	86.0	240.0	163.0	85.0	162.0	37.0	857.
CAN	73.5	146.9	179.5	114.1	64.9	137.7	140.7	
	0.417	0.213	0.487	0.521	0.477	0.429	0.096	
CHI-SQUARE	155.02							
29.	179.0	136.0	324.0	23.0	37.0	138.0	19.0	853.
YOUR	73.1	146.2	178.7	113.6	64.6	137.1	140.1	
	0.888	0.337	0.658	0.073	0.208	0.365	0.049	
CHI-SQUARE	460.80							
30.	60.0	171.0	266.0	75.0	23.0	148.0	110.0	833.
THEY	71.4	142.8	174.5	110.9	63.1	133.9	136.8	
	0.298	0.424	0.540	0.240	0.129	0.392	0.285	
CHI-SQUARE	99.18							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXVII DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF THE CORPUS

RANK WORD	SUBJECTS B	C	D	E	F	G	H	TOTAL
31.	115.0	134.0	42.0	14.0	414.0	72.0	42.0	826.
WE	70.8	141.6	173.1	110.0	62.6	132.8	135.6	
	0.571	0.333	0.085	0.045	2.325	0.191	0.109	
CHI-SQUARE	2277.49							
32.	75.0	346.0	176.0	11.0	28.0	77.0	113.0	816.
HIS	70.0	139.9	171.0	108.6	61.8	131.1	134.0	
	0.372	0.859	0.357	0.035	0.157	0.204	0.293	
CHI-SQUARE	436.17							
33.	136.0	72.0	251.0	120.0	63.0	128.0	46.0	816.
WILL	70.0	139.9	171.0	108.6	61.8	131.1	134.0	
	0.675	0.179	0.510	0.383	0.354	0.339	0.119	
CHI-SQUARE	191.84							
34.	68.0	91.0	278.0	102.0	90.0	118.0	32.0	779.
IF	66.8	133.5	163.2	103.7	59.0	125.2	127.9	
	0.337	0.226	0.564	0.326	0.505	0.312	0.083	
CHI-SQUARE	182.96							
35.	65.0	135.0	142.0	121.0	50.0	135.0	124.0	772.
AN	66.2	132.3	161.7	102.8	58.5	124.1	126.8	
	0.323	0.335	0.288	0.387	0.281	0.357	0.321	
CHI-SQUARE	7.97							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXVII DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF THE CORPUS

RANK WORD	SUBJECTS B	C	D	E	F	G	H	TOTAL
36.	63.0	108.0	193.0	118.0	45.0	124.0	49.0	700.
WHEN	60.0	120.0	146.7	93.2	53.0	112.5	115.0	
	0.313	0.268	0.392	0.377	0.253	0.328	0.127	
CHI-SQUARE	62.82							
37.	86.0	136.0	159.0	61.0	49.0	86.0	104.0	681.
ALL	58.4	116.7	142.7	90.7	51.6	109.5	111.8	
	0.427	0.337	0.323	0.195	0.275	0.228	0.269	
CHI-SQUARE	33.52							
38.	36.0	195.0	135.0	47.0	34.0	96.0	120.0	663.
BUT	56.8	113.6	138.9	88.3	50.2	106.6	108.9	
	0.179	0.484	0.274	0.150	0.191	0.254	0.311	
CHI-SQUARE	92.70							
39.	69.0	64.0	103.0	76.0	87.0	126.0	114.0	639.
THESE	54.8	109.5	133.9	85.1	48.4	102.7	104.9	
	0.342	0.159	0.209	0.243	0.489	0.333	0.295	
CHI-SQUARE	67.56							
40.	68.0	26.0	277.0	65.0	39.0	60.0	46.0	581.
MAY	49.8	99.6	121.7	77.3	44.0	93.4	95.4	
	0.337	0.065	0.562	0.208	0.719	0.159	0.119	
CHI-SQUARE	299.16							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXVII DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF THE CORPUS

RANK WORD	B	C	D	E	F	G	H	TOTAL
41.	30.0	130.0	117.0	77.0	49.0	63.0	103.0	569.
THERE	48.8	97.5	119.2	75.8	43.1	91.5	93.4	
	0.149	0.323	0.238	0.246	0.275	0.167	0.267	
CHI-SQUARE 28.74								
42.	51.0	71.0	102.0	101.0	26.0	80.0	127.0	558.
HAS	47.8	95.6	116.9	74.3	42.3	89.7	91.6	
	0.253	0.176	0.207	0.323	0.146	0.212	0.329	
CHI-SQUARE 39.02								
43.	26.0	422.0	10.0	10.0	4.0	57.0	23.0	552.
I	47.3	94.6	115.6	73.5	41.8	88.7	90.6	
	0.129	1.047	0.020	0.032	0.022	0.151	0.060	
CHI-SQUARE 1389.72								
44.	39.0	62.0	142.0	69.0	46.0	89.0	95.0	542.
OTHER	46.5	92.9	113.6	72.2	41.1	87.1	89.0	
	0.194	0.154	0.288	0.220	0.258	0.236	0.246	
CHI-SQUARE 19.78								
45.	32.0	69.0	165.0	50.0	50.0	77.0	94.0	537.
SOME	46.0	92.0	112.5	71.5	40.7	86.3	88.2	
	0.159	0.171	0.335	0.160	0.281	0.204	0.243	
CHI-SQUARE 44.53								

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXVII DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF THE CORPUS

RANK WORD	B	C	D	E	F	G	H	TOTAL
46.	38.0	75.0	134.0	76.0	26.0	75.0	105.0	529.
MORE	45.3	90.7	110.8	70.4	40.1	85.0	86.9	
	0.189	0.186	0.272	0.243	0.146	0.198	0.272	
CHI-SQUARE 19.09								
47.	16.0	136.0	27.0	28.0	32.0	67.0	212.0	518.
WERE	44.4	88.8	108.5	69.0	39.2	83.3	85.1	
	0.079	0.337	0.055	0.089	0.180	0.177	0.549	
CHI-SQUARE 322.78								
48.	7.0	258.0	34.0	19.0	17.0	59.0	116.0	510.
HAD	43.7	87.4	106.8	67.9	38.6	82.0	83.7	
	0.035	0.640	0.069	0.061	0.095	0.156	0.300	
CHI-SQUARE 479.54								
49.	27.0	102.0	126.0	22.0	13.0	80.0	128.0	498.
THEIR	42.7	85.4	104.3	66.3	37.7	80.0	81.8	
	0.134	0.253	0.256	0.070	0.073	0.212	0.332	
CHI-SQUARE 85.43								
50.	29.0	34.0	141.0	150.0	39.0	63.0	30.0	491.
USED	42.1	84.2	102.9	65.4	37.2	78.9	80.6	
	0.144	0.084	0.286	0.479	0.219	0.180	0.078	
CHI-SQUARE 191.07								

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXVII DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF THE CORPUS

RANK WORD	B	C	D	E	F	G	H	TOTAL
51.	33.0	38.0	115.0	75.0	53.0	84.0	90.0	488.
MANY	41.8	83.6	102.2	65.0	37.0	78.4	80.1	
	0.164	0.094	0.233	0.240	0.298	0.222	0.233	
CHI-SQUARE	38.49							
52.	24.0	110.0	132.0	51.0	41.0	82.0	41.0	481.
SO	41.2	82.4	100.8	64.0	36.4	77.3	79.0	
	0.119	0.273	0.268	0.163	0.230	0.217	0.106	
CHI-SQUARE	47.87							
53.	67.0	36.0	96.0	37.0	87.0	109.0	44.0	476.
EACH	40.8	81.6	99.7	63.4	36.1	76.5	78.2	
	0.332	0.089	0.195	0.118	0.489	0.288	0.114	
CHI-SQUARE	154.13							
54.	33.0	77.0	52.0	75.0	67.0	79.0	82.0	465.
TWO	39.9	79.7	97.4	61.9	35.2	74.7	76.4	
	0.164	0.191	0.106	0.240	0.376	0.209	0.212	
CHI-SQUARE	54.55							
55.	33.0	96.0	70.0	35.0	79.0	88.0	62.0	463.
ABOUT	39.7	79.4	97.0	61.6	35.1	74.4	76.0	
	0.164	0.238	0.142	0.112	0.444	0.233	0.161	
CHI-SQUARE	83.75							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXVII DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF THE CORPUS

RANK WORD	B	C	D	E	F	G	H	TOTAL
56.	51.0	41.0	238.0	38.0	29.0	31.0	11.0	439.
SHOULD	37.6	75.2	92.0	58.4	33.3	70.6	72.1	
	0.253	0.102	0.483	0.121	0.163	0.082	0.028	
CHI-SQUARE	333.82							
57.	39.0	121.0	55.0	12.0	55.0	107.0	39.0	428.
WHAT	36.7	73.4	89.7	57.0	32.4	68.8	70.3	
	0.194	0.300	0.112	0.038	0.309	0.283	0.101	
CHI-SQUARE	130.87							
58.	24.0	59.0	101.0	64.0	31.0	65.0	81.0	425.
THAN	36.4	72.8	89.0	56.6	32.2	68.3	69.8	
	0.119	0.146	0.205	0.204	0.174	0.172	0.210	
CHI-SQUARE	11.46							
59.	41.0	94.0	58.0	42.0	26.0	63.0	100.0	424.
BEEN	36.3	72.7	88.8	56.4	32.1	68.1	69.6	
	0.203	0.233	0.118	0.134	0.146	0.167	0.259	
CHI-SQUARE	36.05							
60.	19.0	105.0	54.0	82.0	11.0	74.0	78.0	423.
INTO	36.3	72.5	88.6	56.3	32.0	68.0	69.5	
	0.094	0.261	0.110	0.262	0.062	0.196	0.202	
CHI-SQUARE	63.42							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXVII DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF THE CORPUS

RANK WORD	B	C	D	E	F	G	H	TOTAL
61.	38.0	97.0	128.0	20.0	35.0	59.0	44.0	421.
THEM	36.1	72.2	88.2	56.0	31.9	67.7	69.1	
	0.189	0.241	0.260	0.064	0.197	0.156	0.114	
CHI-SQUARE	60.34							
62.	46.0	58.0	121.0	55.0	56.0	54.0	27.0	417.
USE	35.7	71.5	87.4	55.5	31.6	67.0	68.5	
	0.228	0.144	0.246	0.176	0.314	0.143	0.070	
CHI-SQUARE	64.96							
63.	40.0	40.0	173.0	72.0	21.0	36.0	29.0	411.
MAKE	35.2	70.5	86.1	54.7	31.1	66.1	67.5	
	0.198	0.099	0.351	0.230	0.118	0.095	0.075	
CHI-SQUARE	145.87							
64.	45.0	82.0	115.0	33.0	45.0	66.0	20.0	406.
DO	34.8	69.6	85.1	54.1	30.8	65.3	66.7	
	0.223	0.203	0.233	0.105	0.253	0.175	0.052	
CHI-SQUARE	63.22							
65.	43.0	110.0	64.0	82.0	7.0	52.0	45.0	403.
UP	34.5	69.1	84.4	53.7	30.5	64.8	66.2	
	0.213	0.273	0.130	0.262	0.039	0.138	0.117	
CHI-SQUARE	73.66							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXVII DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF THE CORPUS

RANK WORD	B	C	D	E	F	G	H	TOTAL
66.	35.0	48.0	103.0	49.0	30.0	65.0	70.0	400.
SUCH	34.3	68.6	83.8	53.3	30.3	64.3	65.7	
	0.174	0.119	0.209	0.157	0.168	0.172	0.181	
CHI-SQUARE	11.21							
67.	45.0	94.0	62.0	61.0	47.0	69.0	22.0	400.
THEN	34.3	68.6	83.8	53.3	30.3	64.3	65.7	
	0.223	0.233	0.126	0.195	0.264	0.183	0.057	
CHI-SQUARE	58.19							
68.	58.0	67.0	109.0	30.0	16.0	52.0	61.0	393.
TIME	33.7	67.4	82.3	52.3	29.8	63.2	64.5	
	0.288	0.166	0.221	0.096	0.090	0.138	0.158	
CHI-SQUARE	44.23							
69.	21.0	77.0	40.0	55.0	4.0	98.0	91.0	386.
ITS	33.1	66.2	80.9	51.4	29.2	62.0	63.4	
	0.104	0.191	0.081	0.176	0.022	0.259	0.236	
CHI-SQUARE	81.76							
70.	36.0	85.0	40.0	29.0	50.0	92.0	37.0	369.
WOULD	31.6	63.3	77.3	49.1	27.9	59.3	60.6	
	0.179	0.211	0.081	0.093	0.281	0.243	0.096	
CHI-SQUARE	78.94							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXVII DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF THE CORPUS

RANK WORD	B	C	D	E	F	G	H	TOTAL
71.	39.0	64.0	57.0	33.0	79.0	75.0	21.0	368.
HOW	31.5	63.1	77.1	49.0	27.9	59.1	60.4	
	0.194	0.159	0.116	0.105	0.444	0.198	0.054	
CHI-SQUARE	135.99							
72.	38.0	7.0	22.0	14.0	228.0	47.0	16.0	366.
NUMBER	31.4	62.7	76.7	48.7	27.7	58.8	60.1	
	0.189	0.017	0.045	0.045	1.280	0.124	0.041	
CHI-SQUARE	1596.28							
73.	24.0	42.0	109.0	98.0	5.0	38.0	47.0	363.
MADE	31.1	62.2	76.1	48.3	27.5	58.3	59.6	
	0.119	0.104	0.221	0.313	0.028	0.101	0.122	
CHI-SQUARE	101.70							
74.	34.0	108.0	67.0	51.0	15.0	39.0	44.0	358.
OUT	30.7	61.4	75.0	47.7	27.1	57.5	58.8	
	0.169	0.268	0.136	0.163	0.084	0.103	0.114	
CHI-SQUARE	51.99							
75.	23.0	47.0	74.0	61.0	8.0	50.0	90.0	353.
MOST	30.3	60.5	74.0	47.0	26.7	56.7	58.0	
	0.114	0.117	0.150	0.195	0.045	0.132	0.233	
CHI-SQUARE	40.56							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXVII DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF THE CORPUS

RANK WORD	B	C	D	E	F	G	H	TOTAL
76.	17.0	64.0	61.0	32.0	37.0	86.0	54.0	351.
ONLY	30.1	60.2	73.5	46.7	26.6	56.4	57.6	
	0.084	0.159	0.124	0.102	0.208	0.228	0.140	
CHI-SQUARE	32.54							
77.	32.0	112.0	68.0	18.0	28.0	52.0	45.0	349.
NO	29.9	59.8	73.1	46.5	26.4	56.1	57.3	
	0.159	0.278	0.138	0.058	0.157	0.138	0.117	
CHI-SQUARE	66.48							
78.	50.0	57.0	73.0	83.0	18.0	44.0	17.0	342.
MUST	29.3	58.6	71.7	45.5	25.9	55.0	56.2	
	0.248	0.141	0.148	0.265	0.101	0.116	0.044	
CHI-SQUARE	77.40							
79.	6.0	30.0	51.0	34.0	5.0	146.0	68.0	340.
WATER	29.1	58.3	71.2	45.3	25.8	54.6	55.8	
	0.030	0.074	0.104	0.109	0.028	0.386	0.176	
CHI-SQUARE	212.75							
80.	23.0	21.0	84.0	68.0	180.0	42.0	56.0	312.
ALSO	26.7	53.5	65.4	41.5	23.6	50.1	51.2	
	0.114	0.052	0.171	0.217	1.011	0.111	0.145	
CHI-SQUARE	1078.83							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXVII DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF THE CORPUS

RANK / WORD	B	C	D	E	F	G	H	TOTAL
81.	27.0	47.0	54.0	41.0	35.0	42.0	62.0	308.
FIRST	26.4	52.8	64.5	41.0	23.3	49.5	50.6	
	0.134	0.117	0.110	0.131	0.197	0.111	0.161	
CHI-SQUARE 11.92								
82.	27.0	48.0	65.0	43.0	20.0	66.0	37.0	306.
VERY	26.2	52.5	64.1	40.7	23.2	49.2	50.2	
	0.134	0.119	0.132	0.137	0.112	0.175	0.096	
CHI-SQUARE 10.22								
83.	41.0	51.0	140.0	33.0	7.0	16.0	15.0	303.
GOOD	26.0	51.9	63.5	40.3	23.0	48.7	49.8	
	0.203	0.127	0.284	0.105	0.039	0.042	0.039	
CHI-SQUARE 159.60								
84.	12.0	160.0	76.0	1.0	12.0	25.0	9.0	295.
HIM	25.3	50.6	61.8	39.3	22.3	47.4	48.4	
	0.060	0.397	0.154	0.003	0.067	0.066	0.023	
CHI-SQUARE 331.87								
85.	25.0	16.0	46.0	40.0	53.0	88.0	18.0	286.
SAME	24.5	49.0	59.9	38.1	21.7	46.0	47.0	
	0.124	0.040	0.093	0.128	0.298	0.233	0.047	
CHI-SQUARE 127.22								

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXVII DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF THE CORPUS

RANK / WORD	B	C	D	E	F	G	H	TOTAL
86.	7.0	83.0	12.0	23.0	45.0	50.0	40.0	260.
COULD	22.3	44.6	54.5	34.6	19.7	41.8	42.7	
	0.035	0.206	0.024	0.073	0.253	0.132	0.104	
CHI-SQUARE 114.95								
87.	52.0	70.0	63.0	11.0	4.0	19.0	41.0	260.
WHO	22.3	44.6	54.5	34.6	19.7	41.8	42.7	
	0.258	0.174	0.128	0.035	0.022	0.050	0.106	
CHI-SQUARE 96.56								
88.	31.0	37.0	59.0	25.0	33.0	47.0	23.0	255.
ANY	21.9	43.7	53.4	33.9	19.3	41.0	41.9	
	0.154	0.092	0.120	0.080	0.185	0.124	0.060	
CHI-SQUARE 26.88								
89.	26.0	30.0	81.0	32.0	10.0	45.0	30.0	254.
BECAUSE	21.8	43.5	53.2	33.8	19.2	40.8	41.7	
	0.129	0.074	0.164	0.102	0.056	0.119	0.078	
CHI-SQUARE 27.79								
90.	14.0	65.0	37.0	22.0	33.0	62.0	20.0	253.
SEE	21.7	43.4	53.0	33.7	19.2	40.7	41.5	
	0.069	0.161	0.075	0.070	0.185	0.164	0.052	
CHI-SQUARE 54.76								

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXVII DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF THE CORPUS

RANK WORD	B	C	D	E	F	G	H	TOTAL
91.	27.0	72.0	52.0	13.0	13.0	38.0	30.0	245.
LIKE	21.0	42.0	51.3	32.6	18.6	39.4	40.2	
	0.134	0.179	0.106	0.042	0.073	0.101	0.078	
CHI-SQUARE	39.27							
92.	25.0	31.0	42.0	28.0	14.0	40.0	59.0	239.
MUCH	20.5	41.0	50.1	31.8	18.1	38.4	39.2	
	0.124	0.077	0.085	0.089	0.079	0.106	0.153	
CHI-SQUARE	16.11							
93.	28.0	40.0	53.0	7.0	8.0	16.0	84.0	236.
PEOPLE	20.2	40.5	49.4	31.4	17.9	37.9	38.8	
	0.139	0.099	0.108	0.022	0.045	0.042	0.218	
CHI-SQUARE	93.18							
94.	17.0	29.0	18.0	44.0	41.0	57.0	28.0	234.
CALLED	20.1	40.1	49.0	31.2	17.7	37.6	38.4	
	0.084	0.072	0.037	0.141	0.230	0.151	0.073	
CHI-SQUARE	71.87							
95.	16.0	22.0	61.0	37.0	20.0	59.0	18.0	233.
PLACE	20.0	39.9	48.8	31.0	17.6	37.4	38.3	
	0.079	0.055	0.124	0.118	0.112	0.156	0.047	
CHI-SQUARE	36.49							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXVII DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF THE CORPUS

RANK WORD	B	C	D	E	F	G	H	TOTAL
96.	10.0	34.0	34.0	54.0	8.0	62.0	30.0	232.
THROUGH	19.9	39.8	48.6	30.9	17.6	37.3	38.1	
	0.050	0.084	0.069	0.173	0.045	0.164	0.078	
CHI-SQUARE	50.75							
97.	39.0	21.0	58.0	60.0	13.0	20.0	21.0	232.
WORK	19.9	39.8	48.6	30.9	17.6	37.3	38.1	
	0.194	0.052	0.118	0.192	0.073	0.053	0.054	
CHI-SQUARE	73.36							
98.	22.0	42.0	27.0	12.0	9.0	29.0	86.0	228.
NEW	19.5	39.1	47.8	30.4	17.3	36.6	37.4	
	0.109	0.104	0.055	0.038	0.051	0.077	0.223	
CHI-SQUARE	89.19							
99.	17.0	14.0	44.0	37.0	10.0	52.0	49.0	223.
SMALL	19.1	38.2	46.7	29.7	16.9	35.8	36.6	
	0.084	0.035	0.089	0.118	0.056	0.138	0.127	
CHI-SQUARE	31.83							
100.	10.0	44.0	45.0	38.0	6.0	33.0	44.0	220.
OVER	18.9	37.7	46.1	29.3	16.7	35.4	36.1	
	0.050	0.109	0.091	0.121	0.034	0.087	0.114	
CHI-SQUARE	16.53							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXVIII DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE EIGHT

RANK WORD	B	C	D	E	F	G	H	TOTAL
1.	0.0	618.0	600.0	336.0	468.0	883.0	954.0	3859.
THE	0.0	528.5	834.4	337.7	516.6	723.5	818.3	
	0.0	7.182	5.252	7.266	6.617	8.913	8.514	
CHI-SQUARE	128.23							
2.	0.0	294.0	289.0	140.0	340.0	420.0	465.0	1949.
OF	0.0	317.4	421.4	170.6	260.9	365.4	413.3	
	0.0	3.417	2.530	3.028	4.807	4.239	4.150	
CHI-SQUARE	87.41							
3.	0.0	249.0	365.0	152.0	130.0	180.0	386.0	1642.
AND	0.0	267.4	355.0	143.7	219.8	307.9	348.2	
	0.0	2.894	3.195	3.287	1.838	1.817	3.445	
CHI-SQUARE	95.92							
4.	0.0	230.0	342.0	138.0	181.0	311.0	234.0	1636.
A	0.0	266.4	353.7	143.2	219.0	306.7	346.9	
	0.0	2.673	2.993	2.984	2.559	3.139	2.088	
CHI-SQUARE	48.97							
5.	0.0	214.0	354.0	127.0	171.0	199.0	261.0	1326.
TO	0.0	215.9	286.7	116.0	177.5	248.6	281.2	
	0.0	2.487	3.098	2.747	2.418	2.009	2.329	
CHI-SQUARE	28.43							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXVIII DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE EIGHT

RANK WORD	B	C	D	E	F	G	H	TOTAL
6.	0.0	161.0	261.0	89.0	150.0	188.0	259.0	1108.
IN	0.0	180.4	239.6	97.0	148.3	207.7	235.0	
	0.0	1.871	2.284	1.925	2.121	1.898	2.311	
CHI-SQUARE	9.02							
7.	0.0	75.0	183.0	145.0	167.0	154.0	146.0	871.
IS	0.0	141.8	188.3	76.2	116.6	163.3	184.7	
	0.0	0.872	1.602	3.136	2.361	1.554	1.303	
CHI-SQUARE	124.15							
8.	0.0	137.0	118.0	12.0	133.0	97.0	70.0	567.
THAT	0.0	92.3	122.6	49.6	75.9	106.3	120.2	
	0.0	1.592	1.033	0.260	1.880	0.979	0.625	
CHI-SQUARE	115.06							
9.	0.0	123.0	136.0	61.0	51.0	101.0	60.0	532.
IT	0.0	86.6	115.0	46.6	71.2	99.7	112.8	
	0.0	1.429	1.190	1.319	0.721	1.019	0.535	
CHI-SQUARE	54.04							
10.	0.0	27.0	183.0	42.0	76.0	66.0	95.0	489.
ARE	0.0	79.6	105.7	42.8	65.5	91.7	103.7	
	0.0	0.314	1.602	0.908	1.075	0.666	0.848	
CHI-SQUARE	100.89							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXVIII DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE EIGHT

RANK WORD	SUBJECTS B	C	D	E	F	G	H	TOTAL
11.	0.0	54.0	175.0	44.0	84.0	43.0	99.0	499.
FOR	0.0	81.3	107.9	43.7	66.8	93.6	105.8	
	0.0	0.628	1.532	0.952	1.188	0.434	0.884	
CHI-SQUARE	83.08							
12.	0.0	83.0	243.0	31.0	63.0	116.0	9.0	545.
YOU	0.0	88.8	117.8	47.7	73.0	102.2	115.6	
	0.0	0.965	2.127	0.670	0.891	1.171	0.080	
CHI-SQUARE	240.65							
13.	0.0	40.0	189.0	37.0	49.0	54.0	49.0	418.
BE	0.0	68.1	90.4	36.6	56.0	78.4	88.6	
	0.0	0.465	1.654	0.800	0.693	0.545	0.437	
CHI-SQUARE	145.36							
14.	0.0	69.0	110.0	39.0	47.0	67.0	73.0	405.
AS	0.0	66.0	87.6	35.4	54.2	75.9	85.9	
	0.0	0.802	0.963	0.843	0.664	0.676	0.651	
CHI-SQUARE	10.19							
15.	0.0	35.0	153.0	34.0	18.0	71.0	58.0	369.
OR	0.0	60.1	79.8	32.3	49.4	69.2	78.2	
	0.0	0.407	1.339	0.735	0.254	0.717	0.518	
CHI-SQUARE	102.99							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXVIII DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE EIGHT

RANK WORD	SUBJECTS B	C	D	E	F	G	H	TOTAL
16.	0.0	51.0	66.0	27.0	42.0	49.0	63.0	298.
WITH	0.0	48.5	64.4	26.1	39.9	55.9	63.2	
	0.0	0.593	0.578	0.584	0.594	0.495	0.562	
CHI-SQUARE	1.15							
17.	0.0	44.0	67.0	31.0	25.0	78.0	95.0	340.
ON	0.0	55.4	73.5	29.8	45.5	63.7	72.1	
	0.0	0.511	0.586	0.670	0.353	0.787	0.848	
CHI-SQUARE	22.67							
18.	0.0	21.0	21.0	21.0	79.0	60.0	61.0	263.
THIS	0.0	42.8	56.9	23.0	35.2	49.3	55.8	
	0.0	0.244	0.184	0.454	1.117	0.606	0.544	
CHI-SQUARE	91.21							
19.	0.0	26.0	40.0	19.0	62.0	35.0	91.0	273.
BY	0.0	44.5	59.0	23.9	36.5	51.2	57.9	
	0.0	0.302	0.350	0.411	0.877	0.353	0.812	
CHI-SQUARE	56.58							
20.	0.0	101.0	14.0	6.0	14.0	38.0	105.0	278.
WAS	0.0	45.3	60.1	24.3	37.2	52.1	59.0	
	0.0	1.174	0.123	0.130	0.198	0.384	0.937	
CHI-SQUARE	172.05							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXVIII DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE EIGHT

RANK WORD	B	C	D	E	F	G	H	TOTAL
	SUBJECTS							
21.	0.0	98.0	19.0	1.0	13.0	46.0	76.0	253.
HE	0.0	41.2	54.7	22.1	33.9	47.4	53.7	
	0.0	1.139	0.166	0.022	0.184	0.464	0.678	
CHI-SQUARE	144.00							
22.	0.0	30.0	34.0	23.0	17.0	69.0	75.0	248.
FROM	0.0	40.4	53.6	21.7	33.2	46.5	52.6	
	0.0	0.349	0.298	0.497	0.240	0.696	0.669	
CHI-SQUARE	38.27							
23.	0.0	44.0	98.0	12.0	42.0	30.0	37.0	263.
HAVE	0.0	42.8	56.9	23.0	35.2	49.3	55.8	
	0.0	0.511	0.858	0.260	0.594	0.303	0.330	
CHI-SQUARE	50.25							
24.	0.0	47.0	57.0	10.0	13.0	74.0	40.0	241.
AT	0.0	39.2	52.1	21.1	32.3	45.2	51.1	
	0.0	0.546	0.499	0.216	0.184	0.747	0.357	
CHI-SQUARE	40.11							
25.	0.0	22.0	42.0	31.0	35.0	59.0	68.0	257.
WHICH	0.0	41.9	55.6	22.5	34.4	48.2	54.5	
	0.0	0.256	0.368	0.670	0.495	0.596	0.607	
CHI-SQUARE	21.73							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXVIII DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE EIGHT

RANK WORD	B	C	D	E	F	G	H	TOTAL
	SUBJECTS							
26.	0.0	33.0	32.0	17.0	54.0	53.0	36.0	225.
ONE	0.0	36.6	48.7	19.7	30.1	42.2	47.7	
	0.0	0.383	0.280	0.368	0.763	0.535	0.321	
CHI-SQUARE	31.01							
27.	0.0	35.0	50.0	10.0	32.0	27.0	38.0	192.
NOT	0.0	31.3	41.5	16.8	25.7	36.0	40.7	
	0.0	0.407	0.438	0.216	0.452	0.273	0.339	
CHI-SQUARE	8.91							
28.	0.0	20.0	60.0	14.0	39.0	55.0	8.0	196.
CAN	0.0	31.9	42.4	17.2	26.2	36.7	41.6	
	0.0	0.232	0.525	0.303	0.551	0.555	0.071	
CHI-SQUARE	54.73							
29.	0.0	36.0	128.0	10.0	14.0	62.0	5.0	255.
YOUR	0.0	41.5	55.1	22.3	34.1	47.8	54.1	
	0.0	0.418	1.120	0.216	0.198	0.626	0.045	
CHI-SQUARE	164.45							
30.	0.0	49.0	66.0	8.0	5.0	36.0	66.0	230.
THEY	0.0	37.5	49.7	20.1	30.8	43.1	48.8	
	0.0	0.569	0.578	0.173	0.071	0.363	0.589	
CHI-SQUARE	45.05							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXVIII DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE EIGHT

RANK / WORD	B	C	D	E	F	G	H	TOTAL
31.	0.0	28.0	14.0	6.0	174.0	23.0	8.0	253.
WE	0.0	41.2	54.7	22.1	33.9	47.4	53.7	
	0.0	0.325	0.123	0.130	2.460	0.232	0.071	
CHI-SQUARE	677.56							
32.	0.0	69.0	26.0	5.0	10.0	36.0	58.0	204.
HIS	0.0	33.2	44.1	17.9	27.3	38.2	43.3	
	0.0	0.802	0.228	0.108	0.141	0.363	0.518	
CHI-SQUARE	71.34							
33.	0.0	19.0	67.0	15.0	24.0	25.0	14.0	164.
WILL	0.0	26.7	35.5	14.4	22.0	30.7	34.8	
	0.0	0.221	0.586	0.324	0.339	0.252	0.125	
CHI-SQUARE	43.98							
34.	0.0	21.0	62.0	8.0	31.0	43.0	10.0	175.
IF	0.0	28.5	37.8	15.3	23.4	32.8	37.1	
	0.0	0.244	0.543	0.173	0.438	0.434	0.089	
CHI-SQUARE	46.31							
35.	0.0	18.0	45.0	27.0	16.0	33.0	37.0	176.
AN	0.0	28.7	38.1	15.4	23.6	33.0	37.3	
	0.0	0.209	0.394	0.584	0.226	0.333	0.330	
CHI-SQUARE	16.40							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXVIII DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE EIGHT

RANK / WORD	B	C	D	E	F	G	H	TOTAL
36.	0.0	20.0	38.0	17.0	18.0	30.0	19.0	142.
WHEN	0.0	23.1	30.7	12.4	19.0	26.6	30.1	
	0.0	0.232	0.333	0.368	0.254	0.303	0.170	
CHI-SQUARE	8.42							
37.	0.0	26.0	30.0	12.0	23.0	17.0	36.0	144.
ALL	0.0	23.5	31.1	12.6	19.3	27.0	30.5	
	0.0	0.302	0.263	0.260	0.325	0.172	0.321	
CHI-SQUARE	5.75							
38.	0.0	32.0	28.0	5.0	17.0	21.0	41.0	144.
BUT	0.0	23.5	31.1	12.6	19.3	27.0	30.5	
	0.0	0.372	0.245	0.108	0.240	0.212	0.366	
CHI-SQUARE	13.21							
39.	0.0	16.0	19.0	8.0	48.0	27.0	30.0	148.
THESE	0.0	24.1	32.0	13.0	19.8	27.7	31.4	
	0.0	0.186	0.166	0.173	0.679	0.273	0.268	
CHI-SQUARE	50.09							
40.	0.0	5.0	64.0	15.0	17.0	8.0	17.0	126.
MAY	0.0	20.5	27.2	11.0	16.9	23.6	26.7	
	0.0	0.058	0.560	0.324	0.240	0.081	0.152	
CHI-SQUARE	76.63							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXVIII   DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE EIGHT

RANK / WORD	B	C	D	E	F	G	H	TOTAL
		SUBJECTS						
41.	0.0	28.0	29.0	3.0	24.0	13.0	45.0	142.
THERE	0.0	23.1	30.7	12.4	19.0	26.6	30.1	
	0.0	0.325	0.254	0.065	0.339	0.131	0.402	
CHI-SQUARE		23.92						
42.	0.0	13.0	17.0	21.0	8.0	16.0	27.0	102.
HAS	0.0	16.6	22.1	8.9	13.7	19.1	21.6	
	0.0	0.151	0.149	0.454	0.113	0.162	0.241	
CHI-SQUARE		22.46						
43.	0.0	90.0	1.0	0.0	3.0	0.0	10.0	104.
I	0.0	16.9	22.5	9.1	13.9	19.5	22.1	
	0.0	1.046	0.009	0.0	0.042	0.0	0.089	
CHI-SQUARE		379.48						
44.	0.0	20.0	42.0	10.0	12.0	24.0	31.0	139.
OTHER	0.0	22.6	30.1	12.2	18.5	26.1	29.5	
	0.0	0.232	0.368	0.216	0.170	0.242	0.277	
CHI-SQUARE		8.03						
45.	0.0	13.0	34.0	4.0	15.0	23.0	33.0	122.
SOME	0.0	19.9	26.4	10.7	16.3	22.9	25.9	
	0.0	0.151	0.298	0.087	0.212	0.232	0.295	
CHI-SQUARE		10.82						

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXVIII   DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE EIGHT

RANK / WORD	B	C	D	E	F	G	H	TOTAL
		SUBJECTS						
46.	0.0	15.0	28.0	7.0	14.0	17.0	27.0	108.
MORE	0.0	17.6	23.4	9.5	14.5	20.2	22.9	
	0.0	0.174	0.245	0.151	0.198	0.172	0.241	
CHI-SQUARE		3.21						
47.	0.0	32.0	8.0	0.0	9.0	28.0	65.0	143.
WERE	0.0	23.3	30.9	12.5	19.1	26.8	30.3	
	0.0	0.372	0.070	0.0	0.127	0.283	0.580	
CHI-SQUARE		77.84						
48.	0.0	39.0	7.0	0.0	9.0	17.0	41.0	113.
HAD	0.0	18.4	24.4	9.9	15.1	21.2	24.0	
	0.0	0.453	0.061	0.0	0.127	0.172	0.366	
CHI-SQUARE		60.80						
49.	0.0	21.0	31.0	4.0	8.0	12.0	44.0	120.
THEIR	0.0	19.5	25.9	10.5	16.1	22.5	25.4	
	0.0	0.244	0.271	0.087	0.113	0.121	0.393	
CHI-SQUARE		27.59						
50.	0.0	6.0	26.0	34.0	21.0	19.0	11.0	117.
USED	0.0	19.1	25.3	10.2	15.7	21.9	24.8	
	0.0	0.070	0.228	0.735	0.297	0.192	0.098	
CHI-SQUARE		74.01						

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXVIII DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE EIGHT

RANK / WORD	B	C	D	E	F	G	H	TOTAL
51.	0.0	5.0	44.0	8.0	12.0	27.0	22.0	118.
MANY	0.0	19.2	25.5	10.3	15.8	22.1	25.0	
	0.0	0.058	0.385	0.173	0.170	0.273	0.196	
CHI-SQUARE	26.79							
52.	0.0	31.0	43.0	3.0	19.0	25.0	16.0	137.
SO	0.0	22.3	29.6	12.0	18.3	25.7	29.1	
	0.0	0.360	0.376	0.065	0.269	0.252	0.143	
CHI-SQUARE	22.07							
53.	0.0	10.0	12.0	11.0	34.0	34.0	21.0	122.
EACH	0.0	19.9	26.4	10.7	16.3	22.9	25.9	
	0.0	0.116	0.105	0.238	0.481	0.343	0.187	
CHI-SQUARE	38.19							
54.	0.0	18.0	7.0	9.0	23.0	29.0	31.0	117.
TWO	0.0	19.1	25.3	10.2	15.7	21.9	24.8	
	0.0	0.209	0.061	0.195	0.325	0.293	0.277	
CHI-SQUARE	20.70							
55.	0.0	10.0	10.0	3.0	14.0	26.0	28.0	91.
ABOUT	0.0	14.8	19.7	8.0	12.2	17.1	19.3	
	0.0	0.116	0.088	0.065	0.198	0.262	0.250	
CHI-SQUARE	18.30							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXVIII DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE EIGHT

RANK / WORD	B	C	D	E	F	G	H	TOTAL
56.	0.0	11.0	74.0	6.0	10.0	6.0	2.0	109.
SHOULD	0.0	17.8	23.6	9.5	14.6	20.4	23.1	
	0.0	0.128	0.648	0.130	0.141	0.061	0.018	
CHI-SQUARE	142.72							
57.	0.0	24.0	20.0	2.0	22.0	42.0	4.0	114.
WHAT	0.0	18.6	24.6	10.0	15.3	21.4	24.2	
	0.0	0.279	0.175	0.043	0.311	0.424	0.036	
CHI-SQUARE	48.56							
58.	0.0	13.0	25.0	12.0	15.0	23.0	14.0	102.
THAN	0.0	16.6	22.1	8.9	13.7	19.1	21.6	
	0.0	0.151	0.219	0.260	0.212	0.232	0.125	
CHI-SQUARE	5.85							
59.	0.0	14.0	10.0	9.0	11.0	8.0	27.0	79.
BEEN	0.0	12.9	17.1	6.9	10.6	14.8	16.8	
	0.0	0.163	0.088	0.195	0.156	0.081	0.241	
CHI-SQUARE	13.08							
60.	0.0	30.0	2.0	8.0	6.0	20.0	30.0	96.
INTO	0.0	15.6	20.8	8.4	12.9	18.0	20.4	
	0.0	0.349	0.018	0.173	0.085	0.202	0.268	
CHI-SQUARE	38.61							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXVIII DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE EIGHT

RANK WORD	B	C	D	E	F	G	H	TOTAL
61.	0.0	21.0	37.0	0.0	16.0	14.0	21.0	109.
THEM	0.0	17.8	23.6	9.5	14.6	20.4	23.1	
	0.0	0.244	0.324	0.0	0.226	0.141	0.187	
CHI-SQUARE	20.15							
62.	0.0	10.0	36.0	12.0	23.0	12.0	4.0	97.
USE	0.0	15.8	21.0	8.5	13.0	18.2	20.6	
	0.0	0.116	0.315	0.260	0.325	0.121	0.036	
CHI-SQUARE	37.52							
63.	0.0	8.0	49.0	11.0	13.0	11.0	15.0	107.
MAKE	0.0	17.4	23.1	9.4	14.3	20.1	22.7	
	0.0	0.093	0.429	0.238	0.184	0.111	0.134	
CHI-SQUARE	41.12							
64.	0.0	12.0	34.0	8.0	15.0	18.0	6.0	93.
DO	0.0	15.1	20.1	8.1	12.4	17.4	19.7	
	0.0	0.139	0.298	0.173	0.212	0.182	0.054	
CHI-SQUARE	20.34							
65.	0.0	17.0	12.0	7.0	2.0	19.0	24.0	81.
UP	0.0	13.2	17.5	7.1	10.8	15.2	17.2	
	0.0	0.198	0.105	0.151	0.028	0.192	0.214	
CHI-SQUARE	13.72							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXVIII DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE EIGHT

RANK WORD	B	C	D	E	F	G	H	TOTAL
66.	0.0	10.0	18.0	12.0	6.0	9.0	22.0	77.
SUCH	0.0	12.5	16.6	6.7	10.3	14.4	16.3	
	0.0	0.116	0.158	0.260	0.085	0.091	0.196	
CHI-SQUARE	10.55							
67.	0.0	18.0	7.0	4.0	10.0	17.0	8.0	109.
THEN	0.0	17.8	23.6	9.5	14.6	20.4	23.1	
	0.0	0.209	0.061	0.087	0.141	0.172	0.071	
CHI-SQUARE	26.77							
68.	0.0	16.0	15.0	0.0	3.0	12.0	20.0	70.
TIME	0.0	11.4	15.1	6.1	9.4	13.1	14.8	
	0.0	0.186	0.131	0.0	0.042	0.121	0.178	
CHI-SQUARE	14.20							
69.	0.0	13.0	4.0	7.0	1.0	26.0	25.0	76.
ITS	0.0	12.4	16.4	6.7	10.2	14.2	16.1	
	0.0	0.151	0.035	0.151	0.014	0.262	0.223	
CHI-SQUARE	32.31							
70.	0.0	20.0	17.0	2.0	13.0	27.0	15.0	94.
WOULD	0.0	15.3	20.3	8.2	12.6	17.6	19.9	
	0.0	0.232	0.149	0.043	0.184	0.273	0.134	
CHI-SQUARE	12.92							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXVIII. DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE EIGHT

RANK WORD	B	C	D	E	F	G	H	TOTAL
71.	0.0	18.0	9.0	3.0	23.0	29.0	3.0	85.
HOW	0.0	13.8	18.4	7.4	11.4	15.9	18.0	
	0.0	0.209	0.079	0.065	0.325	0.293	0.027	
CHI-SQUARE	43.79							
72.	0.0	2.0	6.0	3.0	99.0	5.0	2.0	118.
NUMBER	0.0	19.2	25.5	10.3	15.8	22.1	25.0	
	0.0	0.023	0.053	0.065	1.400	0.050	0.018	
CHI-SQUARE	508.28							
73.	0.0	9.0	32.0	15.0	1.0	18.0	21.0	96.
MADE	0.0	15.6	20.8	8.4	12.9	18.0	20.4	
	0.0	0.105	0.280	0.324	0.014	0.182	0.187	
CHI-SQUARE	25.04							
74.	0.0	24.0	15.0	7.0	7.0	11.0	21.0	85.
OUT	0.0	13.8	18.4	7.4	11.4	15.9	18.0	
	0.0	0.279	0.131	0.151	0.099	0.111	0.187	
CHI-SQUARE	11.81							
75.	0.0	7.0	18.0	7.0	5.0	18.0	21.0	76.
MOST	0.0	12.4	16.4	6.7	10.2	14.2	16.1	
	0.0	0.081	0.158	0.151	0.071	0.182	0.187	
CHI-SQUARE	7.60							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXVIII DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE EIGHT

RANK WORD	B	C	D	E	F	G	H	TOTAL
76.	0.0	17.0	9.0	9.0	27.0	27.0	13.0	102.
ONLY	0.0	16.6	22.1	8.9	13.7	19.1	21.6	
	0.0	0.198	0.079	0.195	0.382	0.273	0.116	
CHI-SQUARE	27.47							
77.	0.0	17.0	8.0	2.0	11.0	11.0	20.0	70.
NO	0.0	11.4	15.1	6.1	9.4	13.1	14.8	
	0.0	0.198	0.070	0.043	0.156	0.111	0.178	
CHI-SQUARE	11.31							
78.	0.0	20.0	19.0	6.0	7.0	12.0	9.0	73.
MUST	0.0	11.9	15.8	6.4	9.8	13.7	15.5	
	0.0	0.232	0.166	0.130	0.099	0.121	0.080	
CHI-SQUARE	9.92							
79.	0.0	6.0	3.0	5.0	0.0	15.0	23.0	52.
WATER	0.0	8.5	11.2	4.6	7.0	9.7	11.0	
	0.0	0.070	0.026	0.108	0.0	0.151	0.205	
CHI-SQUARE	29.60							
80.	0.0	2.0	27.0	7.0	5.0	7.0	12.0	60.
ALSO	0.0	9.8	13.0	5.3	8.0	11.2	12.7	
	0.0	0.023	0.236	0.151	0.071	0.071	0.107	
CHI-SQUARE	24.72							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXVIII DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE EIGHT

RANK WORD	B	C	D	E	F	G	H	TOTAL
81.	0.0	12.0	14.0	5.0	13.0	16.0	13.0	73.
FIRST	0.0	11.9	15.8	6.4	9.8	13.7	15.5	
	0.0	0.139	0.123	0.108	0.184	0.162	0.116	
CHI-SQUARE	2.36							
82.	0.0	7.0	18.0	5.0	10.0	19.0	14.0	73.
VERY	0.0	11.9	15.8	6.4	9.8	13.7	15.5	
	0.0	0.081	0.158	0.108	0.141	0.192	0.125	
CHI-SQUARE	4.83							
83.	0.0	11.0	63.0	6.0	5.0	4.0	4.0	93.
GOOD	0.0	15.1	20.1	8.1	12.4	17.4	19.7	
	0.0	0.128	0.551	0.130	0.071	0.040	0.036	
CHI-SQUARE	120.53							
84.	0.0	31.0	13.0	0.0	3.0	12.0	5.0	64.
HIM	0.0	10.4	13.8	5.6	8.6	12.0	13.6	
	0.0	0.360	0.114	0.0	0.042	0.121	0.045	
CHI-SQUARE	55.31							
85.	0.0	2.0	7.0	5.0	15.0	34.0	7.0	70.
SAME	0.0	11.4	15.1	6.1	9.4	13.1	14.8	
	0.0	0.023	0.061	0.108	0.212	0.343	0.062	
CHI-SQUARE	53.06							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXVIII DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE EIGHT

RANK WORD	B	C	D	E	F	G	H	TOTAL
86.	0.0	18.0	7.0	1.0	16.0	12.0	15.0	69.
COULD	0.0	11.2	14.9	6.0	9.2	12.9	14.6	
	0.0	0.209	0.061	0.022	0.226	0.121	0.134	
CHI-SQUARE	17.51							
87.	0.0	13.0	18.0	2.0	2.0	7.0	22.0	64.
WHO	0.0	10.4	13.8	5.6	8.6	12.0	13.6	
	0.0	0.151	0.158	0.043	0.028	0.071	0.196	
CHI-SQUARE	16.55							
88.	0.0	9.0	17.0	0.0	12.0	16.0	4.0	58.
ANY	0.0	9.4	12.5	5.1	7.8	10.9	12.3	
	0.0	0.105	0.149	0.0	0.170	0.162	0.036	
CHI-SQUARE	17.01							
89.	0.0	4.0	22.0	10.0	4.0	8.0	12.0	60.
BECAUSE	0.0	9.8	13.0	5.3	8.0	11.2	12.7	
	0.0	0.046	0.193	0.216	0.057	0.081	0.107	
CHI-SQUARE	16.99							
90.	0.0	8.0	9.0	2.0	12.0	32.0	4.0	67.
SEE	0.0	10.9	14.5	5.9	9.0	12.6	14.2	
	0.0	0.093	0.079	0.043	0.170	0.323	0.036	
CHI-SQUARE	43.84							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXVIII. DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE EIGHT

RANK WORD	B	C	D	E	F	G	H	TOTAL
91.	0.0	18.0	12.0	3.0	3.0	14.0	14.0	64.
LIKE	0.0	10.4	13.8	5.6	8.6	12.0	13.6	
	0.0	0.209	0.105	0.065	0.042	0.141	0.125	
CHI-SQUARE	10.93							
92.	0.0	6.0	17.0	3.0	5.0	9.0	15.0	55.
MUCH	0.0	9.0	11.9	4.8	7.4	10.3	11.7	
	0.0	0.070	0.149	0.065	0.071	0.091	0.134	
CHI-SQUARE	5.73							
93.	0.0	6.0	20.0	1.0	7.0	6.0	20.0	60.
PEOPLE	0.0	9.8	13.0	5.3	8.0	11.2	12.7	
	0.0	0.070	0.175	0.022	0.099	0.061	0.178	
CHI-SQUARE	15.45							
94.	0.0	5.0	3.0	10.0	14.0	13.0	8.0	53.
CALLED	0.0	8.6	11.5	4.6	7.1	9.9	11.2	
	0.0	0.058	0.026	0.216	0.198	0.131	0.071	
CHI-SQUARE	22.57							
95.	0.0	3.0	25.0	6.0	11.0	20.0	9.0	74.
PLACE	0.0	12.1	16.0	6.5	9.9	13.9	15.7	
	0.0	0.035	0.219	0.130	0.156	0.202	0.080	
CHI-SQUARE	17.57							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXVIII DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE EIGHT

RANK WORD	B	C	D	E	F	G	H	TOTAL
96.	0.0	11.0	5.0	5.0	1.0	13.0	9.0	44.
THROUGH	0.0	7.2	9.5	3.9	5.9	8.2	9.3	
	0.0	0.128	0.044	0.108	0.014	0.131	0.080	
CHI-SQUARE	11.34							
97.	0.0	4.0	8.0	15.0	5.0	3.0	8.0	43.
WORK	0.0	7.0	9.3	3.8	5.8	8.1	9.1	
	0.0	0.046	0.070	0.324	0.071	0.030	0.071	
CHI-SQUARE	38.44							
98.	0.0	6.0	7.0	0.0	5.0	9.0	11.0	38.
NEW	0.0	6.2	8.2	3.3	5.1	7.1	8.1	
	0.0	0.070	0.061	0.0	0.071	0.091	0.098	
CHI-SQUARE	5.08							
99.	0.0	3.0	10.0	3.0	7.0	9.0	16.0	48.
SMALL	0.0	7.8	10.4	4.2	6.4	9.0	10.2	
	0.0	0.035	0.088	0.065	0.099	0.091	0.143	
CHI-SQUARE	6.71							
100.	0.0	8.0	7.0	3.0	4.0	13.0	17.0	52.
OVER	0.0	8.5	11.2	4.6	7.0	9.7	11.0	
	0.0	0.093	0.061	0.065	0.057	0.131	0.152	
CHI-SQUARE	7.73							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXIX DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE NINE

RANK WORD	B	C	D	E	F	G	H	TOTAL
1.	985.0	1575.0	2052.0	2451.0	275.0	1153.0	580.0	9071.
THE	921.3	1706.3	2790.3	1967.0	266.8	906.0	513.2	
	7.889	6.811	5.427	9.195	7.605	9.391	8.339	
CHI-SQUARE 405.18								
2.	352.0	704.0	1077.0	704.0	152.0	556.0	311.0	3857.
OF	391.7	725.5	1186.4	836.4	113.5	385.2	218.2	
	2.819	3.045	2.848	2.641	4.204	4.528	4.472	
CHI-SQUARE 163.93								
3.	288.0	732.0	1288.0	703.0	58.0	246.0	224.0	3537.
AND	359.2	665.3	1088.0	767.0	104.0	353.3	200.1	
	2.307	3.166	3.406	2.637	1.604	2.004	3.221	
CHI-SQUARE 118.72								
4.	313.0	548.0	1027.0	659.0	111.0	306.0	122.0	3086.
A	313.4	580.5	949.3	669.2	90.8	308.2	174.6	
	2.507	2.370	2.716	2.472	3.070	2.492	1.754	
CHI-SQUARE 28.71								
5.	360.0	568.0	1140.0	568.0	72.0	304.0	156.0	3168.
TO	321.8	595.9	974.5	687.0	93.2	316.4	179.2	
	2.883	2.456	3.015	2.131	1.991	2.476	2.243	
CHI-SQUARE 62.89								

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXIX DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE NINE

RANK WORD	B	C	D	E	F	G	H	TOTAL
6.	268.0	450.0	785.0	531.0	62.0	244.0	175.0	2515.
IN	255.4	473.1	773.6	545.4	74.0	251.2	142.3	
	2.147	1.946	2.076	1.992	1.715	1.987	2.516	
CHI-SQUARE 11.95								
7.	203.0	282.0	719.0	651.0	68.0	182.0	102.0	2208.
IS	224.3	415.3	679.2	478.8	65.0	220.5	124.9	
	1.626	1.220	1.902	2.442	1.881	1.482	1.467	
CHI-SQUARE 120.17								
8.	102.0	251.0	296.0	192.0	44.0	148.0	31.0	1064.
THAT	108.1	200.1	327.3	230.7	31.3	106.3	60.2	
	0.817	1.085	0.783	0.720	1.217	1.205	0.446	
CHI-SQUARE 58.45								
9.	109.0	236.0	367.0	276.0	32.0	107.0	54.0	1181.
IT	119.9	222.2	363.3	256.1	34.7	118.0	66.8	
	0.873	1.021	0.971	1.035	0.885	0.871	0.776	
CHI-SQUARE 7.14								
10.	134.0	153.0	517.0	304.0	29.0	79.0	62.0	1278.
ARE	129.8	240.4	393.1	277.1	37.6	127.6	72.3	
	1.073	0.662	1.367	1.140	0.802	0.643	0.891	
CHI-SQUARE 95.53								

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXIX DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE NINE

RANK WORD	B	C	D	E	F	G	H	TOTAL
11.	159.0	188.0	450.0	221.0	28.0	68.0	68.0	1182.
FOR	120.1	222.3	363.6	256.3	34.8	118.1	66.9	
	1.274	0.813	1.190	0.829	0.774	0.554	0.978	
CHI-SQUARE	65.91							
12.	240.0	232.0	390.0	109.0	56.0	139.0	25.0	1191.
YOU	121.0	224.0	366.4	258.3	35.0	119.0	67.4	
	1.922	1.003	1.031	0.409	1.549	1.132	0.359	
CHI-SQUARE	247.80							
13.	118.0	114.0	474.0	288.0	14.0	68.0	26.0	1102.
BE	111.9	207.3	339.0	239.0	32.4	110.1	62.4	
	0.945	0.493	1.254	1.080	0.387	0.554	0.374	
CHI-SQUARE	153.90							
14.	87.0	179.0	296.0	175.0	37.0	78.0	47.0	899.
AS	91.3	169.1	276.5	194.9	26.4	89.8	50.9	
	0.697	0.774	0.783	0.657	1.023	0.635	0.676	
CHI-SQUARE	10.25							
15.	101.0	113.0	483.0	285.0	15.0	56.0	22.0	1075.
OR	109.2	202.2	330.7	233.1	31.6	107.4	60.8	
	0.809	0.489	1.277	1.069	0.415	0.456	0.316	
CHI-SQUARE	179.79							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXIX DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE NINE

RANK WORD	B	C	D	E	F	G	H	TOTAL
16.	70.0	163.0	298.0	204.0	17.0	74.0	44.0	870.
WITH	88.4	163.7	267.6	188.7	25.6	86.9	49.2	
	0.561	0.705	0.788	0.765	0.470	0.603	0.633	
CHI-SQUARE	13.87							
17.	110.0	107.0	209.0	184.0	29.0	93.0	27.0	759.
ON	77.1	142.8	233.5	164.6	22.3	75.8	42.9	
	0.881	0.463	0.553	0.690	0.802	0.757	0.388	
CHI-SQUARE	39.68							
18.	105.0	88.0	127.0	190.0	25.0	78.0	37.0	650.
THIS	66.0	122.3	199.9	141.0	19.1	64.9	36.8	
	0.841	0.381	0.336	0.713	0.691	0.635	0.532	
CHI-SQUARE	80.74							
19.	87.0	106.0	159.0	148.0	14.0	70.0	43.0	627.
BY	63.7	117.9	192.9	136.0	18.4	62.6	35.5	
	0.697	0.458	0.421	0.555	0.387	0.570	0.618	
CHI-SQUARE	20.30							
20.	18.0	222.0	25.0	35.0	24.0	30.0	65.0	419.
WAS	42.6	78.8	128.9	90.9	12.3	41.9	23.7	
	0.144	0.960	0.066	0.131	0.664	0.244	0.935	
CHI-SQUARE	478.70							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXIX DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE NINE

RANK WORD	B	C	D	E	F	G	H	TOTAL
21.	28.0	244.0	168.0	13.0	10.0	45.0	13.0	521.
HE	52.9	98.0	160.3	113.0	15.3	52.0	29.5	
	0.224	1.055	0.444	0.049	0.277	0.367	0.187	
CHI-SQUARE 330.08								
22.	65.0	83.0	155.0	114.0	14.0	78.0	25.0	534.
FROM	54.2	100.4	164.3	115.8	15.7	53.3	30.2	
	0.521	0.359	0.410	0.428	0.387	0.635	0.359	
CHI-SQUARE 18.21								
23.	57.0	110.0	192.0	64.0	16.0	39.0	25.0	503.
HAVE	51.1	94.6	154.7	109.1	14.8	50.2	28.5	
	0.457	0.476	0.508	0.240	0.442	0.318	0.359	
CHI-SQUARE 33.83								
24.	71.0	115.0	150.0	90.0	14.0	66.0	20.0	526.
AT	53.4	98.9	161.8	114.1	15.5	52.5	29.8	
	0.569	0.497	0.397	0.338	0.387	0.538	0.288	
CHI-SQUARE 21.12								
25.	49.0	65.0	156.0	98.0	3.0	65.0	48.0	484.
WHICH	49.2	91.0	148.9	105.0	14.2	48.3	27.4	
	0.392	0.281	0.413	0.368	0.083	0.529	0.690	
CHI-SQUARE 38.38								

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXIX DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE NINE

RANK WORD	B	C	D	E	F	G	H	TOTAL
26.	45.0	87.0	150.0	91.0	18.0	60.0	30.0	481.
ONE	48.9	90.5	148.0	104.3	14.1	48.0	27.2	
	0.360	0.376	0.397	0.341	0.498	0.489	0.431	
CHI-SQUARE 6.47								
27.	32.0	123.0	186.0	67.0	10.0	39.0	35.0	492.
NOT	50.0	92.5	151.3	106.7	14.5	49.1	27.8	
	0.256	0.532	0.492	0.251	0.277	0.318	0.503	
CHI-SQUARE 44.50								
28.	44.0	62.0	180.0	149.0	19.0	52.0	7.0	513.
CAN	52.1	96.5	157.8	111.2	15.1	51.2	29.0	
	0.352	0.268	0.476	0.559	0.525	0.424	0.101	
CHI-SQUARE 47.27								
29.	120.0	87.0	196.0	13.0	16.0	35.0	14.0	481.
YOUR	48.9	90.5	148.0	104.3	14.1	48.0	27.2	
	0.961	0.376	0.518	0.049	0.442	0.285	0.201	
CHI-SQUARE 209.47								
30.	27.0	88.0	200.0	67.0	11.0	34.0	25.0	452.
THEY	45.9	85.0	139.0	98.0	13.3	45.1	25.6	
	0.216	0.381	0.529	0.251	0.304	0.277	0.359	
CHI-SQUARE 47.60								

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXIX DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE NINE

RANK / WORD	SUBJECTS B	C	D	E	F	G	H	TOTAL
31.	40.0	83.0	28.0	8.0	27.0	27.0	6.0	219.
WE	22.2	41.2	67.4	47.5	6.4	21.9	12.4	
	0.320	0.359	0.074	0.030	0.747	0.220	0.086	
CHI-SQUARE	182.54							
32.	34.0	167.0	150.0	6.0	8.0	20.0	22.0	407.
HIS	41.3	76.6	125.2	88.3	12.0	40.7	23.0	
	0.272	0.722	0.397	0.023	0.221	0.163	0.316	
CHI-SQUARE	201.58							
33.	98.0	39.0	184.0	105.0	12.0	44.0	13.0	495.
WILL	50.3	93.1	152.3	107.3	14.6	49.4	28.0	
	0.785	0.169	0.487	0.394	0.332	0.358	0.187	
CHI-SQUARE	92.51							
34.	50.0	56.0	216.0	94.0	12.0	38.0	11.0	477.
IF	48.4	89.7	146.7	103.4	14.0	47.6	27.0	
	0.400	0.242	0.571	0.353	0.332	0.309	0.158	
CHI-SQUARE	58.01							
35.	44.0	91.0	97.0	94.0	11.0	45.0	27.0	409.
AN	41.5	76.9	125.8	88.7	12.0	40.9	23.1	
	0.352	0.394	0.257	0.353	0.364	0.367	0.388	
CHI-SQUARE	10.78							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXIX DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE NINE

RANK / WORD	SUBJECTS B	C	D	E	F	G	H	TOTAL
36.	34.0	67.0	155.0	101.0	15.0	36.0	16.0	424.
WHEN	43.1	79.8	130.4	91.9	12.5	42.3	24.0	
	0.272	0.290	0.410	0.379	0.415	0.293	0.230	
CHI-SQUARE	13.60							
37.	50.0	78.0	129.0	49.0	2.0	39.0	19.0	366.
ALL	37.2	68.8	112.6	79.4	10.8	36.6	20.7	
	0.400	0.337	0.341	0.184	0.055	0.318	0.273	
CHI-SQUARE	27.10							
38.	17.0	116.0	107.0	42.0	3.0	35.0	36.0	356.
BUT	36.2	67.0	109.5	77.2	10.5	35.6	20.1	
	0.136	0.502	0.283	0.158	0.083	0.285	0.518	
CHI-SQUARE	79.98							
39.	45.0	44.0	84.0	68.0	7.0	47.0	24.0	319.
THESE	32.4	60.0	98.1	69.2	9.4	31.9	18.0	
	0.360	0.190	0.222	0.255	0.194	0.383	0.345	
CHI-SQUARE	20.98							
40.	45.0	18.0	213.0	50.0	6.0	18.0	8.0	358.
MAY	36.4	67.3	110.1	77.6	10.5	35.8	20.3	
	0.360	0.078	0.563	0.188	0.166	0.147	0.115	
CHI-SQUARE	162.34							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXIX DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE NINE

RANK WORD	B	C	D	E	F	G	H	TOTAL
41.	22.0	76.0	88.0	74.0	5.0	16.0	20.0	301.
THERE	30.6	56.6	92.6	65.3	8.9	30.1	17.0	
	0.176	0.329	0.233	0.278	0.138	0.130	0.288	
CHI-SQUARE 19.21								
42.	29.0	49.0	85.0	80.0	8.0	30.0	15.0	296.
HAS	30.1	55.7	91.1	64.2	8.7	29.6	16.7	
	0.232	0.212	0.225	0.300	0.221	0.244	0.216	
CHI-SQUARE 5.38								
43.	1.0	215.0	9.0	10.0	0.0	25.0	8.0	268.
I	27.2	50.4	82.4	58.1	7.9	26.8	15.2	
	0.008	0.930	0.024	0.038	0.0	0.204	0.115	
CHI-SQUARE 679.24								
44.	22.0	34.0	100.0	59.0	13.0	27.0	14.0	269.
OTHER	27.3	50.6	82.7	58.3	7.9	26.9	15.2	
	0.176	0.147	0.264	0.221	0.360	0.220	0.201	
CHI-SQUARE 13.46								
45.	21.0	49.0	131.0	46.0	0.0	17.0	19.0	299.
SOME	30.4	56.2	92.0	64.8	8.8	29.9	16.9	
	0.168	0.212	0.346	0.173	0.0	0.138	0.273	
CHI-SQUARE 40.45								

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXIX DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE NINE

RANK WORD	B	C	D	E	F	G	H	TOTAL
46.	21.0	49.0	131.0	46.0	0.0	17.0	19.0	299.
MORE	30.4	56.2	92.0	64.8	8.8	29.9	16.9	
	0.168	0.212	0.346	0.173	0.0	0.138	0.273	
CHI-SQUARE 40.45								
47.	25.0	48.0	106.0	69.0	7.0	26.0	23.0	304.
WERE	30.9	57.2	93.5	65.9	8.9	30.4	17.2	
	0.200	0.208	0.280	0.259	0.194	0.212	0.331	
CHI-SQUARE 7.41								
48.	9.0	74.0	18.0	28.0	5.0	5.0	39.0	178.
HAD	18.1	33.5	54.8	38.6	5.2	17.8	10.1	
	0.072	0.320	0.048	0.105	0.138	0.041	0.561	
CHI-SQUARE 173.46								
49.	3.0	130.0	27.0	19.0	1.0	25.0	14.0	219.
THEIR	22.2	41.2	67.4	47.5	6.4	21.9	12.4	
	0.024	0.562	0.071	0.071	0.028	0.204	0.201	
CHI-SQUARE 254.61								
50.	6.0	60.0	95.0	18.0	4.0	14.0	29.0	226.
USED	23.0	42.5	69.5	49.0	6.6	22.6	12.8	
	0.048	0.259	0.251	0.068	0.111	0.114	0.417	
CHI-SQUARE 73.54								

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXIX DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE NINE

RANK / WORD	B	C	D	E	F	G	H	TOTAL
51.	22.0	26.0	115.0	116.0	9.0	17.0	5.0	310.
MANY	31.5	58.3	95.4	67.2	9.1	31.0	17.5	
	0.176	0.112	0.304	0.435	0.249	0.138	0.072	
CHI-SQUARE	75.47							
52.	15.0	31.0	71.0	67.0	25.0	17.0	19.0	245.
SO	24.9	46.1	75.4	53.1	7.2	24.5	13.9	
	0.120	0.134	0.188	0.251	0.691	0.138	0.273	
CHI-SQUARE	60.85							
53.	18.0	59.0	89.0	48.0	8.0	22.0	5.0	249.
EACH	25.3	46.8	76.6	54.0	7.3	24.9	14.1	
	0.144	0.255	0.235	0.180	0.221	0.179	0.072	
CHI-SQUARE	14.19							
54.	50.0	24.0	84.0	26.0	26.0	34.0	9.0	253.
TWO	25.7	47.6	77.8	54.9	7.4	25.3	14.3	
	0.400	0.104	0.222	0.098	0.719	0.277	0.129	
CHI-SQUARE	101.62							
55.	24.0	38.0	45.0	66.0	11.0	25.0	16.0	225.
ABOUT	22.9	42.3	69.2	48.8	6.6	22.5	12.7	
	0.192	0.164	0.119	0.248	0.304	0.204	0.230	
CHI-SQUARE	19.06							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXIX DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE NINE

RANK / WORD	B	C	D	E	F	G	H	TOTAL
56.	21.0	59.0	60.0	32.0	43.0	33.0	17.0	265.
SHOULD	26.9	49.8	81.5	57.5	7.8	26.5	15.0	
	0.168	0.255	0.159	0.120	1.189	0.269	0.244	
CHI-SQUARE	180.81							
57.	33.0	20.0	164.0	32.0	5.0	10.0	2.0	266.
WHAT	27.0	50.0	81.8	57.7	7.8	26.6	15.1	
	0.264	0.086	0.434	0.120	0.138	0.081	0.029	
CHI-SQUARE	135.99							
58.	18.0	83.0	35.0	10.0	8.0	35.0	5.0	194.
THAN	19.7	36.5	59.7	42.1	5.7	19.4	11.0	
	0.144	0.359	0.093	0.038	0.221	0.285	0.072	
CHI-SQUARE	110.84							
59.	14.0	37.0	76.0	52.0	9.0	21.0	8.0	217.
BEEN	22.0	40.8	66.7	47.1	6.4	21.7	12.3	
	0.112	0.160	0.201	0.195	0.249	0.171	0.115	
CHI-SQUARE	7.68							
60.	19.0	56.0	48.0	33.0	7.0	24.0	10.0	197.
INTO	20.0	37.1	60.6	42.7	5.8	19.7	11.1	
	0.152	0.242	0.127	0.124	0.194	0.195	0.144	
CHI-SQUARE	15.88							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXIX DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE NINE

RANK / WORD	B	C	D	E	F	G	H	TOTAL
61.	14.0	57.0	52.0	74.0	1.0	24.0	15.0	237.
THEM	24.1	44.6	72.9	51.4	7.0	23.7	13.4	
	0.112	0.247	0.138	0.278	0.028	0.195	0.216	
CHI-SQUARE	28.92							
62.	23.0	57.0	91.0	20.0	6.0	15.0	8.0	220.
USE	22.3	41.4	67.7	47.7	6.5	22.0	12.4	
	0.184	0.247	0.241	0.075	0.166	0.122	0.115	
CHI-SQUARE	33.88							
63.	31.0	46.0	85.0	43.0	20.0	21.0	7.0	253.
MAKE	25.7	47.6	77.8	54.9	7.4	25.3	14.3	
	0.248	0.199	0.225	0.161	0.553	0.171	0.101	
CHI-SQUARE	30.02							
64.	27.0	23.0	124.0	61.0	5.0	11.0	4.0	255.
DO	25.9	48.0	78.4	55.3	7.5	25.5	14.4	
	0.216	0.099	0.328	0.229	0.138	0.090	0.058	
CHI-SQUARE	56.69							
65.	27.0	61.0	81.0	25.0	12.0	22.0	6.0	234.
UP	23.8	44.0	72.0	50.7	6.9	23.4	13.2	
	0.216	0.264	0.214	0.094	0.332	0.179	0.086	
CHI-SQUARE	29.03							
66.	26.0	59.0	52.0	75.0	4.0	13.0	11.0	240.
SUCH	24.4	45.1	73.8	52.0	7.1	24.0	13.6	
	0.208	0.255	0.138	0.281	0.111	0.106	0.158	
CHI-SQUARE	27.78							
67.	17.0	34.0	85.0	37.0	7.0	20.0	14.0	214.
THEN	21.7	40.3	65.8	46.4	6.3	21.4	12.1	
	0.136	0.147	0.225	0.139	0.194	0.163	0.201	
CHI-SQUARE	9.96							
68.	37.0	51.0	55.0	57.0	5.0	26.0	7.0	238.
TIME	24.2	44.8	73.2	51.6	7.0	23.8	13.5	
	0.296	0.221	0.145	0.214	0.138	0.212	0.101	
CHI-SQUARE	16.65							
69.	27.0	33.0	94.0	26.0	4.0	18.0	16.0	218.
ITS	22.1	41.0	67.1	47.3	6.4	21.8	12.3	
	0.216	0.143	0.249	0.098	0.111	0.147	0.230	
CHI-SQUARE	25.68							
70.	9.0	44.0	36.0	48.0	0.0	35.0	13.0	185.
WOULD	18.8	34.8	56.9	40.1	5.4	18.5	10.5	
	0.072	0.190	0.095	0.180	0.0	0.285	0.187	
CHI-SQUARE	37.59							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXIX DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE NINE

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXIX — DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE NINE

RANK / WORD	B	C	D	E	F	G	H	TOTAL
	SUBJECTS							
71.	18.0	48.0	23.0	27.0	10.0	30.0	6.0	162.
HOW	16.5	30.5	49.8	35.1	4.8	16.2	9.2	
	0.144	0.208	0.061	0.101	0.277	0.244	0.086	
CHI-SQUARE	45.20							
72.	19.0	40.0	48.0	30.0	34.0	31.0	2.0	204.
NUMBER	20.7	38.4	62.8	44.2	6.0	20.4	11.5	
	0.152	0.173	0.127	0.113	0.940	0.252	0.029	
CHI-SQUARE	152.33							
73.	33.0	3.0	15.0	10.0	18.0	15.0	3.0	98.
MADE	10.0	18.4	30.1	21.3	2.9	9.8	5.5	
	0.264	0.013	0.040	0.038	0.498	0.122	0.043	
CHI-SQUARE	163.07							
74.	16.0	24.0	77.0	83.0	1.0	5.0	4.0	210.
OUT	21.3	39.5	64.6	45.5	6.2	21.0	11.9	
	0.128	0.104	0.204	0.311	0.028	0.041	0.058	
CHI-SQUARE	62.35							
75.	16.0	57.0	52.0	44.0	3.0	19.0	10.0	201.
MOST	20.4	37.8	61.8	43.6	5.9	20.1	11.4	
	0.128	0.247	0.138	0.165	0.083	0.155	0.144	
CHI-SQUARE	13.92							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXIX — DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE NINE

RANK / WORD	B	C	D	E	F	G	H	TOTAL
	SUBJECTS							
76.	10.0	31.0	56.0	54.0	2.0	9.0	11.0	173.
ONLY	17.6	32.5	53.2	37.5	5.1	17.3	9.8	
	0.080	0.134	0.148	0.203	0.055	0.073	0.158	
CHI-SQUARE	16.72							
77.	12.0	34.0	52.0	23.0	4.0	21.0	15.0	161.
NO	16.4	30.3	49.5	34.9	4.7	16.1	9.1	
	0.096	0.147	0.138	0.086	0.111	0.171	0.216	
CHI-SQUARE	11.23							
78.	20.0	76.0	59.0	15.0	2.0	15.0	4.0	192.
MUST	19.5	36.1	59.1	41.6	5.6	19.2	10.9	
	0.160	0.329	0.156	0.056	0.055	0.122	0.058	
CHI-SQUARE	68.70							
79.	36.0	29.0	54.0	77.0	0.0	18.0	2.0	216.
WATER	21.9	40.6	66.4	46.8	6.4	21.6	12.2	
	0.288	0.125	0.143	0.289	0.0	0.147	0.029	
CHI-SQUARE	49.59							
80.	1.0	15.0	48.0	29.0	5.0	94.0	6.0	198.
ALSO	20.1	37.2	60.9	42.9	5.8	19.8	11.2	
	0.008	0.065	0.127	0.109	0.138	0.766	0.086	
CHI-SQUARE	319.80							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXIX DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE NINE

RANK WORD	B	C	D	E	F	G	H	TOTAL
81.	13.0	17.0	57.0	61.0	0.0	15.0	11.0	174.
FIRST	17.7	32.7	53.5	37.7	5.1	17.4	9.8	
	0.104	0.074	0.151	0.229	0.0	0.122	0.158	
CHI-SQUARE	28.95							
82.	16.0	22.0	40.0	36.0	4.0	12.0	12.0	142.
VERY	14.4	26.7	43.7	30.8	4.2	14.2	8.0	
	0.128	0.095	0.106	0.135	0.111	0.098	0.173	
CHI-SQUARE	4.50							
83.	10.0	34.0	47.0	38.0	0.0	22.0	6.0	157.
GOOD	15.9	29.5	48.3	34.0	4.6	15.7	8.9	
	0.080	0.147	0.124	0.143	0.0	0.179	0.086	
CHI-SQUARE	11.49							
84.	23.0	31.0	77.0	27.0	0.0	4.0	6.0	168.
HIM	17.1	31.6	51.7	36.4	4.9	16.8	9.5	
	0.184	0.134	0.204	0.101	0.0	0.033	0.086	
CHI-SQUARE	32.90							
85.	7.0	76.0	63.0	1.0	1.0	10.0	2.0	160.
SAME	16.3	30.1	49.2	34.7	4.7	16.0	9.1	
	0.056	0.329	0.167	0.004	0.028	0.081	0.029	
CHI-SQUARE	122.51							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXIX DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE NINE

RANK WORD	B	C	D	E	F	G	H	TOTAL
86.	14.0	11.0	39.0	35.0	8.0	25.0	5.0	137.
COULD	13.9	25.8	42.1	29.7	4.0	13.7	7.8	
	0.112	0.048	0.103	0.131	0.221	0.204	0.072	
CHI-SQUARE	23.89							
87.	0.0	41.0	5.0	22.0	8.0	11.0	8.0	95.
WHO	9.6	17.9	29.2	20.6	2.8	9.5	5.4	
	0.0	0.177	0.013	0.083	0.221	0.090	0.115	
CHI-SQUARE	70.98							
88.	22.0	41.0	45.0	9.0	2.0	4.0	9.0	132.
ANY	13.4	24.8	40.6	28.6	3.9	13.2	7.5	
	0.176	0.177	0.119	0.034	0.055	0.033	0.129	
CHI-SQUARE	37.59							
89.	14.0	24.0	42.0	25.0	4.0	13.0	4.0	126.
BECAUSE	12.8	23.7	38.8	27.3	3.7	12.6	7.1	
	0.112	0.104	0.111	0.094	0.111	0.106	0.058	
CHI-SQUARE	2.00							
90.	16.0	23.0	59.0	22.0	3.0	13.0	8.0	144.
SEE	14.6	27.1	44.3	31.2	4.2	14.4	8.1	
	0.128	0.099	0.156	0.083	0.083	0.106	0.115	
CHI-SQUARE	8.85							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXIX DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE NINE

RANK WORD	B	C	D	E	F	G	H	TOTAL
				SUBJECTS				
91.	11.0	39.0	28.0	20.0	5.0	20.0	10.0	133.
LIKE	13.5	25.0	40.9	28.8	3.9	13.3	7.5	
	0.088	0.169	0.074	0.075	0.138	0.163	0.144	
CHI-SQUARE	19.58							
92.	9.0	40.0	40.0	10.0	4.0	17.0	9.0	129.
MUCH	13.1	24.3	39.7	28.0	3.8	12.9	7.3	
	0.072	0.173	0.106	0.038	0.111	0.138	0.129	
CHI-SQUARE	24.76							
93.	12.0	21.0	25.0	25.0	0.0	16.0	10.0	116.
PEOPLE	11.8	21.8	35.7	25.2	3.4	11.6	6.6	
	0.096	0.091	0.066	0.094	0.0	0.130	0.144	
CHI-SQUARE	10.13							
94.	6.0	27.0	33.0	6.0	1.0	2.0	29.0	104.
CALLED	10.6	19.6	32.0	22.6	3.1	10.4	5.9	
	0.048	0.117	0.087	0.023	0.028	0.016	0.417	
CHI-SQUARE	115.95							
95.	7.0	18.0	15.0	34.0	13.0	17.0	7.0	111.
PLACE	11.3	20.9	34.1	24.1	3.3	11.1	6.3	
	0.056	0.078	0.040	0.128	0.360	0.138	0.101	
CHI-SQUARE	49.11							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXIX DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE NINE

RANK WORD	B	C	D	E	F	G	H	TOTAL
				SUBJECTS				
96.	13.0	14.0	36.0	31.0	2.0	25.0	2.0	123.
THROUGH	12.5	23.1	37.8	26.7	3.6	12.3	7.0	
	0.104	0.061	0.095	0.116	0.055	0.204	0.029	
CHI-SQUARE	21.84							
97.	5.0	17.0	29.0	49.0	1.0	24.0	11.0	136.
WORK	13.8	25.6	41.8	29.5	4.0	13.6	7.7	
	0.040	0.074	0.077	0.184	0.028	0.195	0.158	
CHI-SQUARE	37.00							
98.	30.0	16.0	50.0	45.0	6.0	9.0	7.0	163.
NEW	16.6	30.7	50.1	35.3	4.8	16.3	9.2	
	0.340	0.069	0.132	0.169	0.166	0.073	0.101	
CHI-SQUARE	24.66							
99.	11.0	21.0	20.0	12.0	2.0	8.0	11.0	85.
SMALL	8.6	16.0	26.1	18.4	2.5	8.5	4.8	
	0.088	0.091	0.053	0.045	0.055	0.065	0.158	
CHI-SQUARE	14.01							
100.	6.0	9.0	34.0	34.0	1.0	20.0	13.0	117.
OVER	11.9	22.0	36.0	25.4	3.4	11.7	6.6	
	0.048	0.039	0.090	0.128	0.028	0.163	0.187	
CHI-SQUARE	27.44							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXX DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE TEN

RANK WORD	SUBJECTS B	C	D	E	F	G	H	TOTAL
1.	478.0	606.0	0.0	0.0	520.0	1230.0	1755.0	4589.
THE	591.9	661.7	0.0	0.0	549.3	1205.6	1580.4	
	6.248	7.085	0.0	0.0	7.324	7.893	8.591	
CHI-SQUARE	47.96							
2.	254.0	190.0	0.0	0.0	264.0	659.0	1005.0	2373.
OF	306.1	342.2	0.0	0.0	284.0	623.4	817.3	
	3.320	2.221	0.0	0.0	3.718	4.229	4.920	
CHI-SQUARE	123.12							
3.	183.0	276.0	0.0	0.0	99.0	320.0	580.0	1458.
AND	188.1	210.2	0.0	0.0	174.5	383.0	502.1	
	2.392	3.227	0.0	0.0	1.394	2.054	2.839	
CHI-SQUARE	75.84							
4.	189.0	214.0	0.0	0.0	233.0	416.0	347.0	1399.
A	180.5	201.7	0.0	0.0	167.5	367.5	481.8	
	2.470	2.502	0.0	0.0	3.282	2.670	1.699	
CHI-SQUARE	70.91							
5.	239.0	193.0	0.0	0.0	178.0	356.0	461.0	1427.
TO	184.1	205.8	0.0	0.0	170.8	374.9	491.5	
	3.124	2.257	0.0	0.0	2.507	2.285	2.257	
CHI-SQUARE	20.33							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXX DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE TEN

RANK WORD	SUBJECTS B	C	D	E	F	G	H	TOTAL
6.	143.0	166.0	0.0	0.0	146.0	367.0	577.0	1399.
IN	180.5	201.7	0.0	0.0	167.5	367.5	481.8	
	1.869	1.941	0.0	0.0	2.056	2.355	2.825	
CHI-SQUARE	35.66							
7.	146.0	48.0	0.0	0.0	191.0	258.0	192.0	836.
IS	107.8	120.5	0.0	0.0	100.1	219.6	287.9	
	1.908	0.561	0.0	0.0	2.690	1.656	0.940	
CHI-SQUARE	178.45							
8.	98.0	94.0	0.0	0.0	130.0	180.0	133.0	635.
THAT	81.9	91.6	0.0	0.0	76.0	166.8	218.7	
	1.281	1.099	0.0	0.0	1.831	1.155	0.651	
CHI-SQUARE	76.20							
9.	67.0	119.0	0.0	0.0	57.0	144.0	118.0	505.
IT	65.1	72.8	0.0	0.0	60.4	132.7	173.9	
	0.876	1.391	0.0	0.0	0.803	0.924	0.578	
CHI-SQUARE	48.48							
10.	79.0	16.0	0.0	0.0	76.0	143.0	84.0	398.
ARE	51.3	57.4	0.0	0.0	47.6	104.6	137.1	
	1.033	0.187	0.0	0.0	1.070	0.918	0.411	
CHI-SQUARE	96.32							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXX DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE TEN

RANK WORD	B	C	D	E	F	G	H	TOTAL
11.	99.0	75.0	0.0	0.0	33.0	94.0	159.0	460.
FOR	59.3	66.3	0.0	0.0	55.1	120.8	158.4	
	1.294	0.877	0.0	0.0	0.465	0.603	0.778	
CHI-SQUARE 42.46								
12.	126.0	50.0	0.0	0.0	68.0	91.0	19.0	354.
YOU	45.7	51.0	0.0	0.0	42.4	93.0	121.9	
	1.647	0.585	0.0	0.0	0.958	0.584	0.093	
CHI-SQUARE 243.79								
13.	73.0	31.0	0.0	0.0	58.0	112.0	77.0	351.
BE	45.3	50.6	0.0	0.0	42.0	92.2	120.9	
	0.954	0.352	0.0	0.0	0.817	0.719	0.377	
CHI-SQUARE 50.84								
14.	48.0	61.0	0.0	0.0	64.0	129.0	183.0	485.
AS	62.6	69.9	0.0	0.0	58.1	127.4	167.0	
	0.627	0.713	0.0	0.0	0.901	0.828	0.896	
CHI-SQUARE 6.68								
15.	37.0	18.0	0.0	0.0	38.0	71.0	42.0	206.
OR	26.6	29.7	0.0	0.0	24.7	54.1	70.9	
	0.484	0.210	0.0	0.0	0.535	0.456	0.206	
CHI-SQUARE 33.00								

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXX DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE TEN

RANK WORD	B	C	D	E	F	G	H	TOTAL
16.	28.0	70.0	0.0	0.0	25.0	101.0	93.0	317.
WITH	40.9	45.7	0.0	0.0	37.9	83.3	109.2	
	0.366	0.818	0.0	0.0	0.352	0.648	0.455	
CHI-SQUARE 27.55								
17.	55.0	59.0	0.0	0.0	43.0	73.0	130.0	360.
ON	46.4	51.9	0.0	0.0	43.1	94.6	124.0	
	0.719	0.690	0.0	0.0	0.606	0.468	0.636	
CHI-SQUARE 7.76								
18.	58.0	27.0	0.0	0.0	61.0	92.0	140.0	378.
THIS	48.8	54.5	0.0	0.0	45.2	99.3	130.2	
	0.758	0.316	0.0	0.0	0.859	0.590	0.685	
CHI-SQUARE 22.40								
19.	39.0	34.0	0.0	0.0	36.0	93.0	144.0	346.
BY	44.6	49.9	0.0	0.0	41.4	90.9	119.2	
	0.510	0.398	0.0	0.0	0.507	0.597	0.705	
CHI-SQUARE 11.71								
20.	6.0	127.0	0.0	0.0	21.0	68.0	191.0	413.
WAS	53.3	59.6	0.0	0.0	49.4	108.5	142.2	
	0.078	1.485	0.0	0.0	0.296	0.436	0.935	
CHI-SQUARE 166.53								

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXX DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE TEN

RANK WORD	B	C	D	E	F	G	H	TOTAL
21.	51.0	178.0	0.0	0.0	19.0	46.0	19.0	313.
HE	40.4	45.1	0.0	0.0	37.5	82.2	107.8	
	0.667	2.081	0.0	0.0	0.268	0.295	0.093	
CHI-SQUARE	492.15							
22.	25.0	35.0	0.0	0.0	24.0	95.0	103.0	282.
FROM	36.4	40.7	0.0	0.0	33.8	74.1	97.1	
	0.327	0.409	0.0	0.0	0.338	0.610	0.504	
CHI-SQUARE	13.43							
23.	88.0	26.0	0.0	0.0	56.0	69.0	52.0	291.
HAVE	37.5	42.0	0.0	0.0	34.8	76.5	100.2	
	1.150	0.304	0.0	0.0	0.789	0.443	0.255	
CHI-SQUARE	110.71							
24.	41.0	62.0	0.0	0.0	22.0	82.0	80.0	287.
AT	37.0	41.4	0.0	0.0	34.4	75.4	98.8	
	0.536	0.725	0.0	0.0	0.310	0.526	0.392	
CHI-SQUARE	19.31							
25.	21.0	19.0	0.0	0.0	46.0	76.0	89.0	251.
WHICH	32.4	36.2	0.0	0.0	30.0	65.9	86.4	
	0.274	0.222	0.0	0.0	0.648	0.488	0.436	
CHI-SQUARE	22.25							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXX DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE TEN

RANK WORD	B	C	D	E	F	G	H	TOTAL
26.	23.0	21.0	0.0	0.0	49.0	67.0	67.0	227.
ONE	29.3	32.7	0.0	0.0	27.2	59.6	78.2	
	0.301	0.246	0.0	0.0	0.690	0.430	0.328	
CHI-SQUARE	25.60							
27.	31.0	43.0	0.0	0.0	46.0	56.0	29.0	205.
NOT	26.4	29.6	0.0	0.0	24.5	53.9	70.6	
	0.405	0.503	0.0	0.0	0.648	0.359	0.142	
CHI-SQUARE	50.27							
28.	40.0	4.0	0.0	0.0	27.0	55.0	22.0	148.
CAN	19.1	21.3	0.0	0.0	17.7	38.9	51.0	
	0.523	0.047	0.0	0.0	0.380	0.353	0.108	
CHI-SQUARE	65.01							
29.	59.0	13.0	0.0	0.0	7.0	41.0	0.0	120.
YOUR	15.5	17.3	0.0	0.0	14.4	31.5	41.3	
	0.771	0.152	0.0	0.0	0.099	0.263	0.0	
CHI-SQUARE	171.39							
30.	33.0	34.0	0.0	0.0	7.0	78.0	19.0	171.
THEY	22.1	24.7	0.0	0.0	20.5	44.9	58.9	
	0.431	0.398	0.0	0.0	0.099	0.501	0.093	
CHI-SQUARE	69.21							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXX DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE TEN

RANK WORD	B	C	D	E	F	G	H	TOTAL
31.	75.0	23.0	0.0	0.0	213.0	22.0	28.0	361.
WE	46.6	52.1	0.0	0.0	43.2	94.8	124.3	
	0.980	0.269	0.0	0.0	3.000	0.141	0.137	
CHI-SQUARE 831.29								
32.	41.0	110.0	0.0	0.0	10.0	21.0	33.0	215.
HIS	27.7	31.0	0.0	0.0	25.7	56.5	74.0	
	0.536	1.286	0.0	0.0	0.141	0.135	0.162	
CHI-SQUARE 262.31								
33.	38.0	14.0	0.0	0.0	27.0	59.0	19.0	157.
WILL	20.3	22.6	0.0	0.0	18.8	41.2	54.1	
	0.497	0.164	0.0	0.0	0.380	0.379	0.093	
CHI-SQUARE 52.82								
34.	18.0	14.0	0.0	0.0	47.0	37.0	11.0	127.
IF	16.4	18.3	0.0	0.0	15.2	33.4	43.7	
	0.235	0.164	0.0	0.0	0.662	0.237	0.054	
CHI-SQUARE 92.59								
35.	21.0	26.0	0.0	0.0	23.0	57.0	60.0	187.
AN	24.1	27.0	0.0	0.0	22.4	49.1	64.4	
	0.274	0.304	0.0	0.0	0.324	0.366	0.294	
CHI-SQUARE 2.02								

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXX DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE TEN

RANK WORD	B	C	D	E	F	G	H	TOTAL
36.	29.0	21.0	0.0	0.0	12.0	58.0	14.0	134.
WHEN	17.3	19.3	0.0	0.0	16.0	35.2	46.1	
	0.379	0.246	0.0	0.0	0.169	0.372	0.069	
CHI-SQUARE 46.26								
37.	36.0	32.0	0.0	0.0	24.0	30.0	49.0	171.
ALL	22.1	24.7	0.0	0.0	20.5	44.9	58.9	
	0.471	0.374	0.0	0.0	0.338	0.193	0.240	
CHI-SQUARE 18.23								
38.	19.0	47.0	0.0	0.0	14.0	40.0	43.0	163.
BUT	21.0	23.5	0.0	0.0	19.5	42.8	56.1	
	0.248	0.550	0.0	0.0	0.197	0.257	0.210	
CHI-SQUARE 28.50								
39.	24.0	4.0	0.0	0.0	32.0	52.0	60.0	172.
THESE	22.2	24.8	0.0	0.0	20.6	45.2	59.2	
	0.314	0.047	0.0	0.0	0.451	0.334	0.294	
CHI-SQUARE 24.96								
40.	23.0	3.0	0.0	0.0	16.0	34.0	21.0	97.
MAY	12.5	14.0	0.0	0.0	11.6	25.5	33.4	
	0.301	0.035	0.0	0.0	0.225	0.218	0.103	
CHI-SQUARE 26.53								

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXX DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE TEN

RANK WORD	B	C	D	E	F	G	H	TOTAL
41.	8.0	26.0	0.0	0.0	20.0	34.0	38.0	197.
THERE	25.4	28.4	0.0	0.0	23.6	51.8	67.8	
	0.105	0.304	0.0	0.0	0.282	0.218	0.186	
CHI-SQUARE	31.90							
42.	22.0	9.0	0.0	0.0	10.0	34.0	85.0	160.
HAS	20.6	23.1	0.0	0.0	19.2	42.0	55.1	
	0.286	0.105	0.0	0.0	0.141	0.218	0.416	
CHI-SQUARE	30.80							
43.	25.0	117.0	0.0	0.0	1.0	32.0	5.0	180.
I	23.2	26.0	0.0	0.0	21.5	47.3	62.0	
	0.327	1.368	0.0	0.0	0.014	0.205	0.024	
CHI-SQUARE	396.43							
44.	17.0	8.0	0.0	0.0	21.0	38.0	50.0	134.
OTHER	17.3	19.3	0.0	0.0	16.0	35.2	46.1	
	0.222	0.094	0.0	0.0	0.296	0.244	0.245	
CHI-SQUARE	8.72							
45.	11.0	7.0	0.0	0.0	19.0	37.0	42.0	116.
SOME	15.0	16.7	0.0	0.0	13.9	30.5	40.0	
	0.144	0.082	0.0	0.0	0.268	0.237	0.206	
CHI-SQUARE	10.09							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXX DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE TEN

RANK WORD	B	C	D	E	F	G	H	TOTAL
46.	13.0	12.0	0.0	0.0	5.0	32.0	55.0	117.
MORE	15.1	16.9	0.0	0.0	14.0	30.7	40.3	
	0.170	0.140	0.0	0.0	0.070	0.205	0.269	
CHI-SQUARE	12.90							
47.	7.0	30.0	0.0	0.0	18.0	34.0	108.0	197.
WERE	25.4	28.4	0.0	0.0	23.6	51.8	67.8	
	0.091	0.351	0.0	0.0	0.254	0.218	0.529	
CHI-SQUARE	44.60							
48.	4.0	89.0	0.0	0.0	7.0	17.0	61.0	178.
HAD	23.0	25.7	0.0	0.0	21.3	46.8	61.3	
	0.052	1.041	0.0	0.0	0.099	0.109	0.299	
CHI-SQUARE	200.48							
49.	21.0	21.0	0.0	0.0	1.0	54.0	55.0	152.
THEIR	19.6	21.9	0.0	0.0	18.2	39.9	52.3	
	0.274	0.246	0.0	0.0	0.014	0.347	0.269	
CHI-SQUARE	21.48							
50.	7.0	2.0	0.0	0.0	9.0	32.0	14.0	64.
USED	8.3	9.2	0.0	0.0	7.7	16.8	22.0	
	0.091	0.023	0.0	0.0	0.127	0.205	0.069	
CHI-SQUARE	22.74							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXX DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE TEN

RANK WORD	B	C	D	E	F	G	H	TOTAL
	SUBJECTS							
51.	18.0	2.0	0.0	0.0	16.0	40.0	49.0	125.
MANY	16.1	18.0	0.0	0.0	15.0	32.8	43.0	
	0.235	0.023	0.0	0.0	0.225	0.257	0.240	
CHI-SQUARE	16.92							
52.	6.0	20.0	0.0	0.0	14.0	35.0	20.0	95.
SO	12.3	13.7	0.0	0.0	11.4	25.0	32.7	
	0.078	0.234	0.0	0.0	0.197	0.225	0.098	
CHI-SQUARE	15.68							
53.	17.0	2.0	0.0	0.0	27.0	41.0	14.0	101.
EACH	13.0	14.6	0.0	0.0	12.1	26.5	34.8	
	0.222	0.023	0.0	0.0	0.380	0.263	0.069	
CHI-SQUARE	50.74							
54.	9.0	21.0	0.0	0.0	33.0	25.0	35.0	123.
TWO	15.9	17.7	0.0	0.0	14.7	32.3	42.4	
	0.118	0.246	0.0	0.0	0.465	0.160	0.171	
CHI-SQUARE	29.19							
55.	12.0	27.0	0.0	0.0	22.0	29.0	17.0	107.
ABOUT	13.8	15.4	0.0	0.0	12.8	28.1	36.9	
	0.157	0.316	0.0	0.0	0.310	0.186	0.083	
CHI-SQUARE	26.23							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXX DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE TEN

RANK WORD	B	C	D	E	F	G	H	TOTAL
	SUBJECTS							
56.	18.0	10.0	0.0	0.0	14.0	15.0	7.0	64.
SHOULD	8.3	9.2	0.0	0.0	7.7	16.8	22.0	
	0.235	0.117	0.0	0.0	0.197	0.096	0.034	
CHI-SQUARE	27.27							
57.	21.0	14.0	0.0	0.0	25.0	30.0	30.0	120.
WHAT	15.5	17.3	0.0	0.0	14.4	31.5	41.3	
	0.274	0.164	0.0	0.0	0.352	0.193	0.147	
CHI-SQUARE	13.65							
58.	10.0	9.0	0.0	0.0	7.0	21.0	59.0	106.
THAN	13.7	15.3	0.0	0.0	12.7	27.8	36.5	
	0.131	0.105	0.0	0.0	0.099	0.135	0.289	
CHI-SQUARE	21.66							
59.	22.0	24.0	0.0	0.0	8.0	31.0	63.0	148.
BEEN	19.1	21.3	0.0	0.0	17.7	38.9	51.0	
	0.288	0.281	0.0	0.0	0.113	0.199	0.308	
CHI-SQUARE	10.54							
60.	5.0	18.0	0.0	0.0	4.0	30.0	33.0	90.
INTO	11.6	13.0	0.0	0.0	10.8	23.6	31.0	
	0.065	0.210	0.0	0.0	0.056	0.193	0.162	
CHI-SQUARE	11.80							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXX DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE TEN

RANK WORD	B	C	D	E	F	G	H	TOTAL
61.	15.0	19.0	0.0	0.0	13.0	30.0	15.0	92.
THEM	11.9	13.3	0.0	0.0	11.0	24.2	31.7	
	0.196	0.222	0.0	0.0	0.183	0.193	0.073	
CHI-SQUARE	13.86							
62.	15.0	2.0	0.0	0.0	13.0	21.0	16.0	67.
USE	8.6	9.7	0.0	0.0	8.0	17.6	23.1	
	0.196	0.023	0.0	0.0	0.183	0.135	0.078	
CHI-SQUARE	16.67							
63.	13.0	9.0	0.0	0.0	3.0	14.0	10.0	49.
MAKE	6.3	7.1	0.0	0.0	5.9	12.9	16.9	
	0.170	0.105	0.0	0.0	0.042	0.090	0.049	
CHI-SQUARE	11.89							
64.	18.0	9.0	0.0	0.0	18.0	26.0	8.0	79.
DO	10.2	11.4	0.0	0.0	9.5	20.8	27.2	
	0.235	0.105	0.0	0.0	0.254	0.167	0.039	
CHI-SQUARE	29.09							
65.	17.0	34.0	0.0	0.0	1.0	20.0	10.0	82.
UP	10.6	11.8	0.0	0.0	9.8	21.5	28.2	
	0.222	0.398	0.0	0.0	0.014	0.128	0.049	
CHI-SQUARE	65.30							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXX DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE TEN

RANK WORD	B	C	D	E	F	G	H	TOTAL
66.	18.0	4.0	0.0	0.0	17.0	36.0	34.0	109.
SUCH	14.1	15.7	0.0	0.0	13.0	28.6	37.5	
	0.235	0.047	0.0	0.0	0.239	0.231	0.166	
CHI-SQUARE	13.26							
67.	8.0	25.0	0.0	0.0	32.0	26.0	7.0	98.
THEN	12.6	14.1	0.0	0.0	11.7	25.7	33.8	
	0.105	0.292	0.0	0.0	0.451	0.167	0.034	
CHI-SQUARE	66.29							
68.	31.0	18.0	0.0	0.0	9.0	22.0	25.0	105.
TIME	13.5	15.1	0.0	0.0	12.6	27.6	36.2	
	0.405	0.210	0.0	0.0	0.127	0.141	0.122	
CHI-SQUARE	28.63							
69.	12.0	20.0	0.0	0.0	3.0	37.0	53.0	125.
ITS	16.1	18.0	0.0	0.0	15.0	32.8	43.0	
	0.157	0.234	0.0	0.0	0.042	0.237	0.259	
CHI-SQUARE	13.66							
70.	18.0	17.0	0.0	0.0	27.0	35.0	16.0	113.
WOULD	14.6	16.3	0.0	0.0	13.5	29.7	38.9	
	0.235	0.199	0.0	0.0	0.380	0.225	0.078	
CHI-SQUARE	28.70							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXX DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE TEN

RANK WORD	B	C	D	E	F	G	H	TOTAL
71.	20.0	6.0	0.0	0.0	22.0	15.0	16.0	79.
HOW	10.2	11.4	0.0	0.0	9.5	20.8	27.2	
	0.261	0.070	0.0	0.0	0.310	0.096	0.078	
CHI-SQUARE	34.85							
72.	4.0	1.0	0.0	0.0	110.0	26.0	10.0	152.
NUMBER	19.6	21.9	0.0	0.0	18.2	39.9	52.3	
	0.052	0.012	0.0	0.0	1.549	0.167	0.049	
CHI-SQUARE	534.74							
73.	8.0	9.0	0.0	0.0	3.0	15.0	22.0	57.
MADE	7.4	8.2	0.0	0.0	6.8	15.0	19.6	
	0.105	0.105	0.0	0.0	0.042	0.096	0.108	
CHI-SQUARE	2.56							
74.	18.0	27.0	0.0	0.0	5.0	9.0	13.0	72.
OUT	9.3	10.4	0.0	0.0	8.6	18.9	24.8	
	0.235	0.316	0.0	0.0	0.070	0.058	0.064	
CHI-SQUARE	47.10							
75.	13.0	9.0	0.0	0.0	1.0	23.0	58.0	104.
MOST	13.4	15.0	0.0	0.0	12.4	27.3	35.8	
	0.170	0.105	0.0	0.0	0.014	0.148	0.284	
CHI-SQUARE	27.36							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXX DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE TEN

RANK WORD	B	C	D	E	F	G	H	TOTAL
76.	5.0	13.0	0.0	0.0	6.0	38.0	26.0	88.
ONLY	11.4	12.7	0.0	0.0	10.5	23.1	30.3	
	0.065	0.152	0.0	0.0	0.085	0.244	0.127	
CHI-SQUARE	15.70							
77.	11.0	18.0	0.0	0.0	14.0	25.0	20.0	89.
NO	11.5	12.8	0.0	0.0	10.7	23.4	30.7	
	0.144	0.210	0.0	0.0	0.197	0.160	0.098	
CHI-SQUARE	6.96							
78.	14.0	8.0	0.0	0.0	11.0	14.0	6.0	53.
MUST	6.8	7.6	0.0	0.0	6.3	13.9	18.3	
	0.183	0.094	0.0	0.0	0.155	0.090	0.029	
CHI-SQUARE	19.17							
79.	5.0	9.0	0.0	0.0	0.0	37.0	39.0	90.
WATER	11.6	13.0	0.0	0.0	10.8	23.6	31.0	
	0.065	0.105	0.0	0.0	0.0	0.237	0.191	
CHI-SQUARE	25.37							
80.	10.0	2.0	0.0	0.0	13.0	20.0	33.0	78.
ALSO	10.1	11.2	0.0	0.0	9.3	20.5	26.9	
	0.131	0.023	0.0	0.0	0.183	0.128	0.162	
CHI-SQUARE	10.45							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXX. DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE TEN

RANK WORD	B	C	D	E	F	G	H	TOTAL
81.	11.0	13.0	0.0	0.0	18.0	14.0	37.0	93.
FIRST	12.0	13.4	0.0	0.0	11.1	24.4	32.0	
	0.144	0.152	0.0	0.0	0.254	0.090	0.181	
CHI-SQUARE 9.56								
82.	17.0	7.0	0.0	0.0	10.0	25.0	17.0	76.
VERY	9.8	11.0	0.0	0.0	9.1	20.0	26.2	
	0.222	0.082	0.0	0.0	0.141	0.160	0.083	
CHI-SQUARE 11.29								
83.	18.0	9.0	0.0	0.0	2.0	8.0	5.0	42.
GOOD	5.4	6.1	0.0	0.0	5.0	11.0	14.5	
	0.235	0.105	0.0	0.0	0.028	0.051	0.024	
CHI-SQUARE 39.50								
84.	5.0	53.0	0.0	0.0	8.0	3.0	2.0	71.
HIM	9.2	10.2	0.0	0.0	8.5	18.7	24.5	
	0.065	0.620	0.0	0.0	0.113	0.019	0.010	
CHI-SQUARE 214.28								
85.	11.0	3.0	0.0	0.0	30.0	29.0	6.0	79.
SAME	10.2	11.4	0.0	0.0	9.5	20.8	27.2	
	0.144	0.035	0.0	0.0	0.423	0.186	0.029	
CHI-SQUARE 70.68								

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXX DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE TEN

RANK WORD	B	C	D	E	F	G	H	TOTAL
86.	7.0	24.0	0.0	0.0	21.0	27.0	17.0	96.
COULD	12.4	13.8	0.0	0.0	11.5	25.2	33.1	
	0.091	0.281	0.0	0.0	0.296	0.173	0.083	
CHI-SQUARE 25.59								
87.	30.0	16.0	0.0	0.0	0.0	8.0	10.0	64.
WHO	8.3	9.2	0.0	0.0	7.7	16.8	22.0	
	0.392	0.187	0.0	0.0	0.0	0.051	0.049	
CHI-SQUARE 81.10								
88.	17.0	4.0	0.0	0.0	17.0	18.0	15.0	71.
ANY	9.2	10.2	0.0	0.0	8.5	18.7	24.5	
	0.222	0.047	0.0	0.0	0.239	0.116	0.073	
CHI-SQUARE 22.70								
89.	10.0	3.0	0.0	0.0	3.0	24.0	10.0	50.
BECAUSE	6.4	7.2	0.0	0.0	6.0	13.1	17.2	
	0.131	0.035	0.0	0.0	0.042	0.154	0.049	
CHI-SQUARE 17.91								
90.	3.0	18.0	0.0	0.0	16.0	10.0	6.0	53.
SEE	6.8	7.6	0.0	0.0	6.3	13.9	18.3	
	0.039	0.210	0.0	0.0	0.225	0.064	0.029	
CHI-SQUARE 40.22								

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXX DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE TEN

RANK WORD	B	C	D	E	F	G	H	TOTAL
91.	18.0	14.0	0.0	0.0	6.0	7.0	7.0	52.
LIKE	6.7	7.5	0.0	0.0	6.2	13.7	17.9	
	0.235	0.164	0.0	0.0	0.085	0.045	0.034	
CHI-SQUARE	34.55							
92.	13.0	4.0	0.0	0.0	2.0	15.0	34.0	68.
MUCH	8.8	9.8	0.0	0.0	8.1	17.9	23.4	
	0.170	0.047	0.0	0.0	0.028	0.096	0.166	
CHI-SQUARE	15.35							
93.	22.0	7.0	0.0	0.0	0.0	8.0	35.0	72.
PEOPLE	9.3	10.4	0.0	0.0	8.6	18.9	24.8	
	0.288	0.082	0.0	0.0	0.0	0.051	0.171	
CHI-SQUARE	37.62							
94.	10.0	6.0	0.0	0.0	14.0	27.0	13.0	70.
CALLED	9.0	10.1	0.0	0.0	8.4	18.4	24.1	
	0.131	0.070	0.0	0.0	0.197	0.173	0.064	
CHI-SQUARE	14.68							
95.	3.0	5.0	0.0	0.0	7.0	14.0	7.0	36.
PLACE	4.6	5.2	0.0	0.0	4.3	9.5	12.4	
	0.039	0.058	0.0	0.0	0.099	0.090	0.034	
CHI-SQUARE	6.80							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

TABLE XXXX DISTRIBUTION OF OCCURRENCE OF THE 100 MOST FREQUENT WORD TYPES ACROSS THE SUBJECT AREAS OF GRADE TEN

RANK WORD	B	C	D	E	F	G	H	TOTAL
96.	5.0	6.0	0.0	0.0	6.0	25.0	10.0	52.
THROUGH	6.7	7.5	0.0	0.0	6.2	13.7	17.9	
	0.065	0.070	0.0	0.0	0.085	0.160	0.049	
CHI-SQUARE	13.65							
97.	9.0	1.0	0.0	0.0	2.0	8.0	6.0	26.
WORK	3.4	3.7	0.0	0.0	3.1	6.8	9.0	
	0.118	0.012	0.0	0.0	0.028	0.051	0.029	
CHI-SQUARE	13.09							
98.	12.0	15.0	0.0	0.0	2.0	12.0	64.0	105.
NEW	13.5	15.1	0.0	0.0	12.6	27.6	36.2	
	0.157	0.175	0.0	0.0	0.028	0.077	0.313	
CHI-SQUARE	39.30							
99.	11.0	2.0	0.0	0.0	2.0	23.0	20.0	58.
SMALL	7.5	8.4	0.0	0.0	6.9	15.2	20.0	
	0.144	0.023	0.0	0.0	0.028	0.148	0.098	
CHI-SQUARE	13.97							
100.	6.0	10.0	0.0	0.0	2.0	9.0	18.0	45.
OVER	5.8	6.5	0.0	0.0	5.4	11.8	15.5	
	0.078	0.117	0.0	0.0	0.026	0.058	0.088	
CHI-SQUARE	5.11							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF WORDS IN SUBJECT

APPENDIX J

CHI SQUARE RESULTS OF DISTRIBUTION OF SELECTED SENTENCE LENGTHS

TABLE XXXXI — DISTRIBUTION OF OCCURRENCE OF FIVE SELECTED SENTENCE LENGTHS ACROSS THE GRADE LEVELS OF THE CORPUS

RANK LENGTH	GRADES 8	9	10	TOTAL
1.	105.0	294.0	182.0	581.
10	126.0	305.6	149.4	
	3.693	4.262	5.396	
CHI-SQUARE		11.02		
2.	115.0	227.0	128.0	470.
20	101.9	247.2	120.9	
	4.045	3.291	3.795	
CHI-SQUARE		3.76		
3.	50.0	96.0	46.0	192.
30	41.6	101.0	49.4	
	1.759	1.392	1.364	
CHI-SQUARE		2.16		
4.	16.0	24.0	8.0	48.
40	10.4	25.2	12.3	
	0.563	0.348	0.237	
CHI-SQUARE		4.60		
5.	29.0	81.0	39.0	149.
50+	32.3	78.4	38.3	
	1.020	1.174	1.156	
CHI-SQUARE		0.44		

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF SENT-LENGTH IN GRADE

TABLE XXXXII — DISTRIBUTION OF OCCURRENCE OF FIVE SELECTED SENTENCE LENGTHS ACROSS THE SUBJECT AREAS OF THE CORPUS

RANK LENGTH	SUBJECTS B	C	D	E	F	G	H	TOTAL
1.	52.0	87.0	107.0	92.0	67.0	92.0	84.0	581.
10	50.2	101.6	118.1	83.1	51.7	90.5	85.6	
	4.586	3.793	4.014	4.904	5.736	4.503	4.346	
CHI-SQUARE		8.72						
2.	38.0	57.0	107.0	66.0	49.0	73.0	81.0	471.
20	40.7	82.4	95.8	67.4	41.9	73.4	69.4	
	3.351	2.485	4.014	3.518	4.195	3.573	4.190	
CHI-SQUARE		12.47						
3.	18.0	35.0	38.0	31.0	6.0	25.0	39.0	192.
30	16.6	33.6	39.0	27.5	17.1	29.9	28.3	
	1.587	1.526	1.425	1.652	0.514	1.224	2.018	
CHI-SQUARE		12.72						
4.	4.0	14.0	3.0	5.0	1.0	6.0	15.0	48.
40	4.2	8.4	9.8	6.9	4.3	7.5	7.1	
	0.353	0.610	0.113	0.267	0.086	0.294	0.776	
CHI-SQUARE		20.61						
5.	7.0	70.0	15.0	9.0	4.0	25.0	20.0	150.
50+	13.0	26.2	30.5	21.5	13.4	23.4	22.1	
	0.617	3.051	0.563	0.480	0.342	1.224	1.035	
CHI-SQUARE		97.71						

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF SENT-LENGTH IN SUBJECT

TABLE XXXXIII DISTRIBUTION OF OCCURRENCE OF FIVE SELECTED SENTENCE LENGTHS ACROSS THE SUBJECT AREAS OF GRADE EIGHT

RANK LENGTH		SUBJECTS B	C	D	E	F	G	H	TOTAL
1.		0.0	19.0	23.0	6.0	17.0	26.0	14.0	105.
	10	0.0	18.1	21.7	9.8	15.0	21.4	19.1	
		0.0	3.885	3.912	2.273	4.187	4.483	2.713	
	CHI-SQUARE			4.16					
2.		0.0	16.0	26.0	7.0	25.0	18.0	23.0	115.
	20	0.0	19.8	23.8	10.7	16.4	23.5	20.9	
		0.0	3.272	4.422	2.652	6.158	3.103	4.457	
	CHI-SQUARE			8.16					
3.		0.0	11.0	8.0	6.0	5.0	7.0	13.0	50.
	30	0.0	8.6	10.3	4.6	7.1	10.2	9.1	
		0.0	2.249	1.361	2.273	1.232	1.207	2.519	
	CHI-SQUARE			4.94					
4.		0.0	4.0	4.0	1.0	1.0	1.0	11.0	22.
	40	0.0	3.8	4.6	2.0	3.1	4.5	4.0	
		0.0	0.818	0.680	0.379	0.246	0.172	2.132	
	CHI-SQUARE			17.08					
5.		0.0	12.0	2.0	0.0	2.0	4.0	8.0	28.
	50+	0.0	4.8	5.8	2.6	4.0	5.7	5.1	
		0.0	2.454	0.340	0.0	0.493	0.690	1.550	
	CHI-SQUARE			18.99					

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF SENT-LENGTH IN SUBJECT

TABLE XXXXIV DISTRIBUTION OF OCCURRENCE OF FIVE SELECTED SENTENCE LENGTHS ACROSS THE SUBJECT AREAS OF GRADE TEN

RANK	LENGTH	B	C	D	E	F	G	H	TOTAL
		S U B J E C T S							
1.		23.0	26.0	0.0	0.0	37.0	40.0	57.0	183.
	10	22.9	31.0	0.0	0.0	27.7	42.2	59.1	
		5.437	4.545	0.0	0.0	7.241	5.141	5.234	
	CHI-SQUARE	4.11							
2.		23.0	13.0	0.0	0.0	13.0	32.0	48.0	129.
	20	16.2	21.9	0.0	0.0	19.5	29.8	41.6	
		5.437	2.273	0.0	0.0	2.544	4.113	4.408	
	CHI-SQUARE	9.81							
3.		7.0	7.0	0.0	0.0	1.0	11.0	21.0	47.
	30	5.9	8.0	0.0	0.0	7.1	10.8	15.2	
		1.655	1.224	0.0	0.0	0.196	1.414	1.928	
	CHI-SQUARE	7.83							
4.		1.0	2.0	0.0	0.0	1.0	2.0	3.0	9.
	40	1.1	1.5	0.0	0.0	1.4	2.1	2.9	
		0.236	0.350	0.0	0.0	0.196	0.257	0.275	
	CHI-SQUARE	0.26							
5.		1.0	12.0	0.0	0.0	2.0	17.0	7.0	39.
	50+	4.9	6.6	0.0	0.0	5.9	9.0	12.6	
		0.236	2.098	0.0	0.0	0.391	2.185	0.643	
	CHI-SQUARE	19.67							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF SENT-LENGTH IN SUBJECT

TABLE XXXXV DISTRIBUTION OF OCCURRENCE OF FIVE SELECTED SENTENCE LENGTHS ACROSS THE SUBJECT AREAS OF GRADE NINE

RANK	LENGTH	B	C	D	E	F	G	H	TOTAL
		S U B J E C T S							
1.		29.0	42.0	84.0	86.0	14.0	26.0	13.0	294.
	10	30.3	52.6	88.6	68.7	10.7	29.2	14.0	
		4.079	3.406	4.042	5.335	5.578	3.796	3.963	
	CHI-SQUARE	8.20							
2.		15.0	28.0	81.0	59.0	11.0	23.0	10.0	227.
	20	23.4	40.6	68.4	53.0	8.3	22.5	10.8	
		2.110	2.271	3.898	3.660	4.382	3.358	3.049	
	CHI-SQUARE	10.88							
3.		11.0	17.0	30.0	25.0	1.0	7.0	5.0	96.
	30	9.9	17.2	28.9	22.4	3.5	9.5	4.6	
		1.547	1.379	1.444	1.551	0.398	1.022	1.524	
	CHI-SQUARE	2.95							
4.		3.0	8.0	3.0	4.0	1.0	3.0	2.0	24.
	40	2.5	4.3	7.2	5.6	0.9	2.4	1.1	
		0.422	0.649	0.144	0.248	0.398	0.438	0.610	
	CHI-SQUARE	7.08							
5.		6.0	45.0	13.0	9.0	0.0	4.0	4.0	81.
	50+	8.3	14.5	24.4	18.9	2.9	8.0	3.9	
		0.844	3.650	0.626	0.558	0.0	0.584	1.220	
	CHI-SQUARE	80.52							

THE THREE LINES OF FIGURES FOR EACH ENTRY REPRESENT:
FREQUENCY
EXPECTED FREQUENCY
RATIO AS %, OF FREQ. TO TOTAL NO. OF SENT-LENGTH IN SUBJECT

APPENDIX K

WORD FREQUENCY DIAGRAMS (GRAPHS)

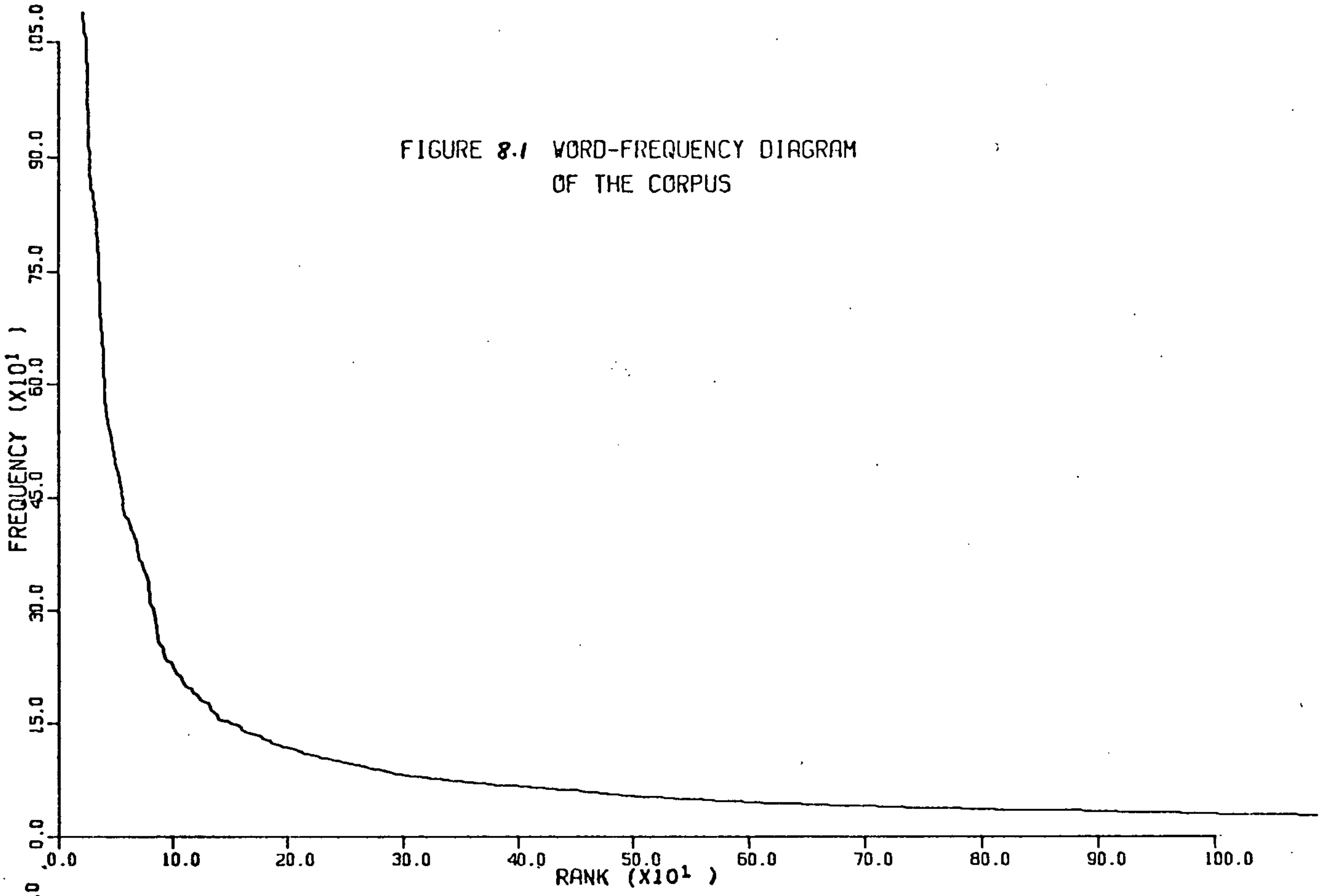


FIGURE 8.1 WORD-FREQUENCY DIAGRAM OF THE CORPUS

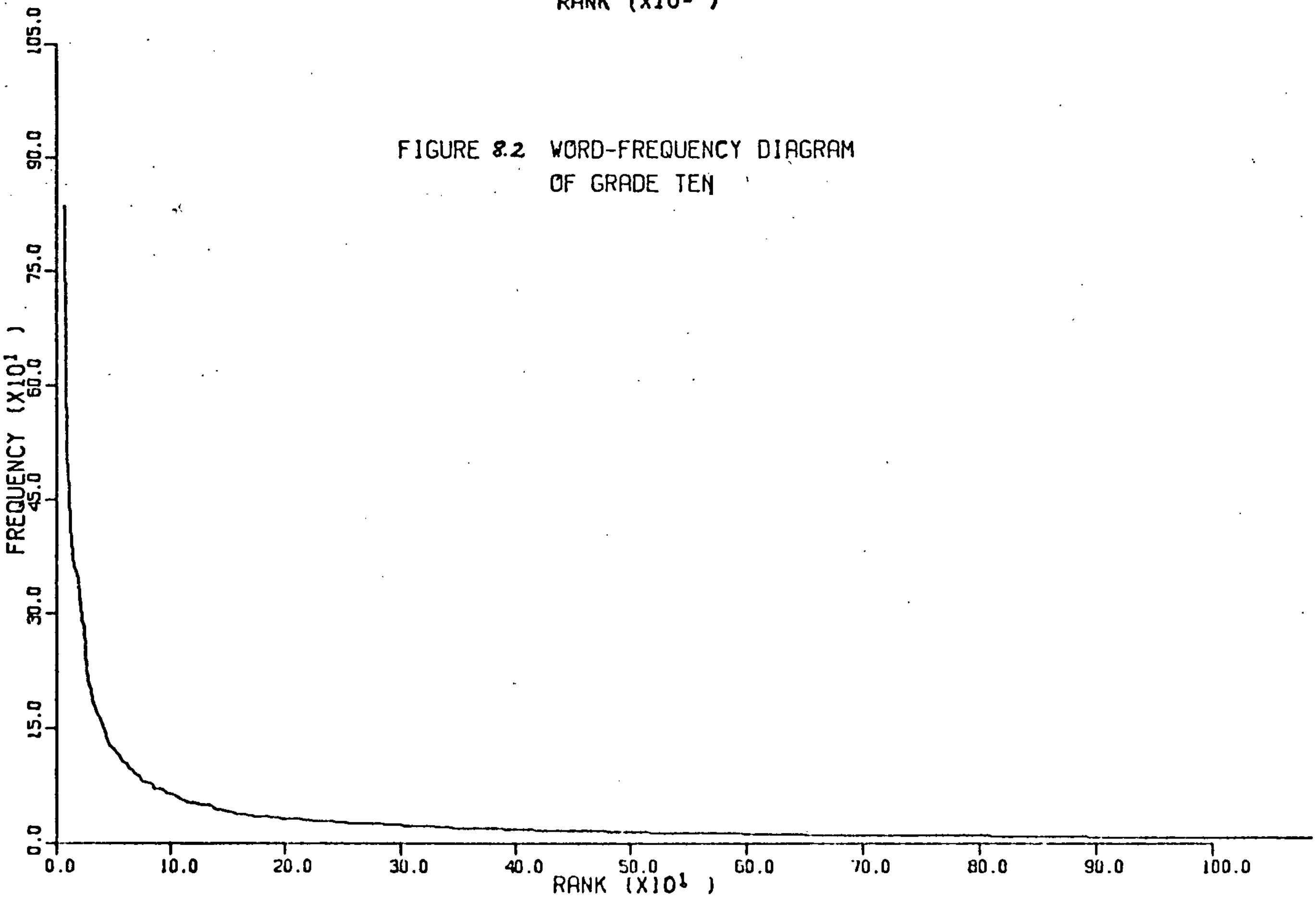


FIGURE 8.2 WORD-FREQUENCY DIAGRAM OF GRADE TEN

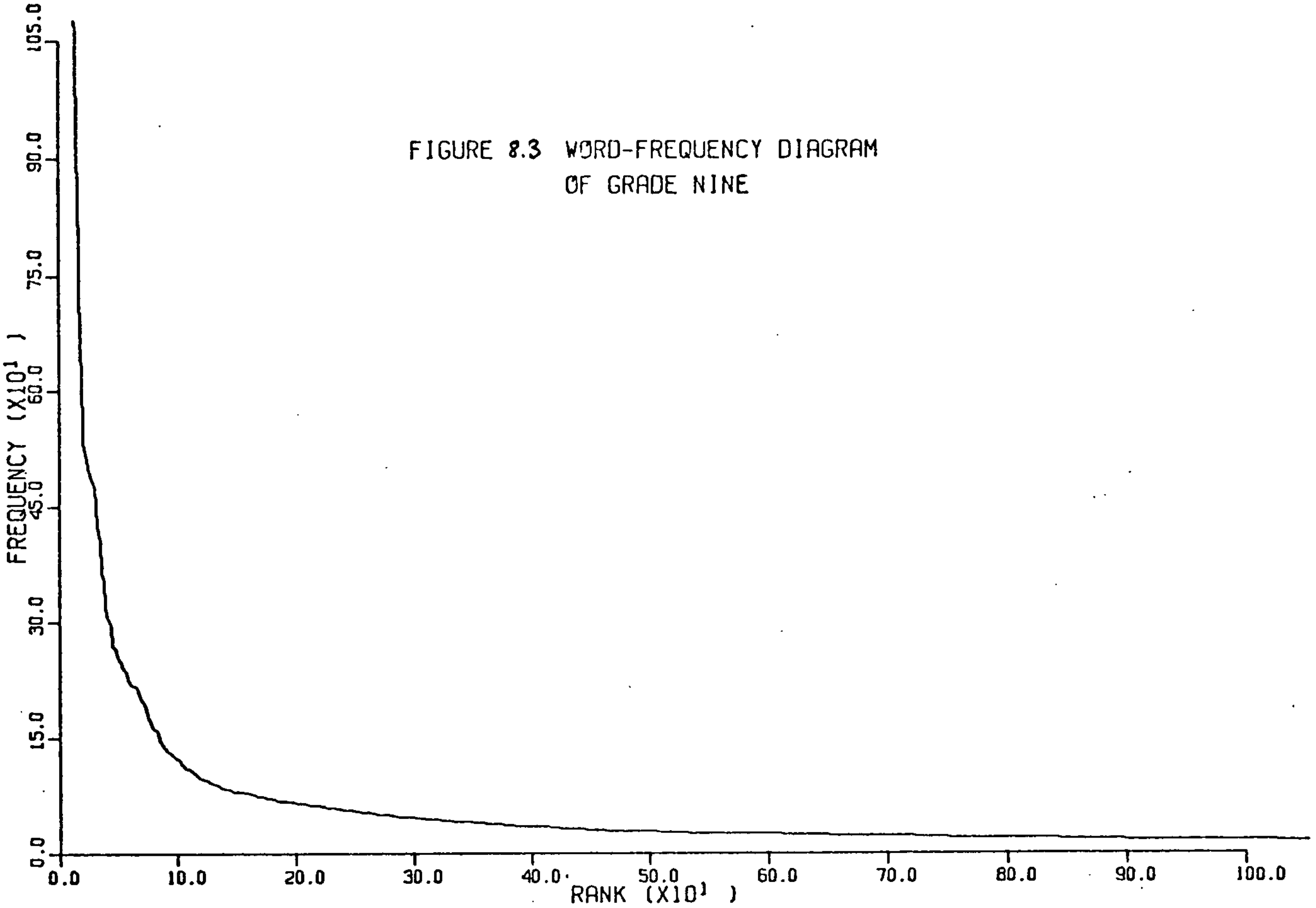


FIGURE 8.3 WORD-FREQUENCY DIAGRAM
OF GRADE NINE
FREQUENCY (X10¹)
RANK (X10¹)
0.0
15.0
30.0
45.0
60.0
75.0
90.0
105.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0

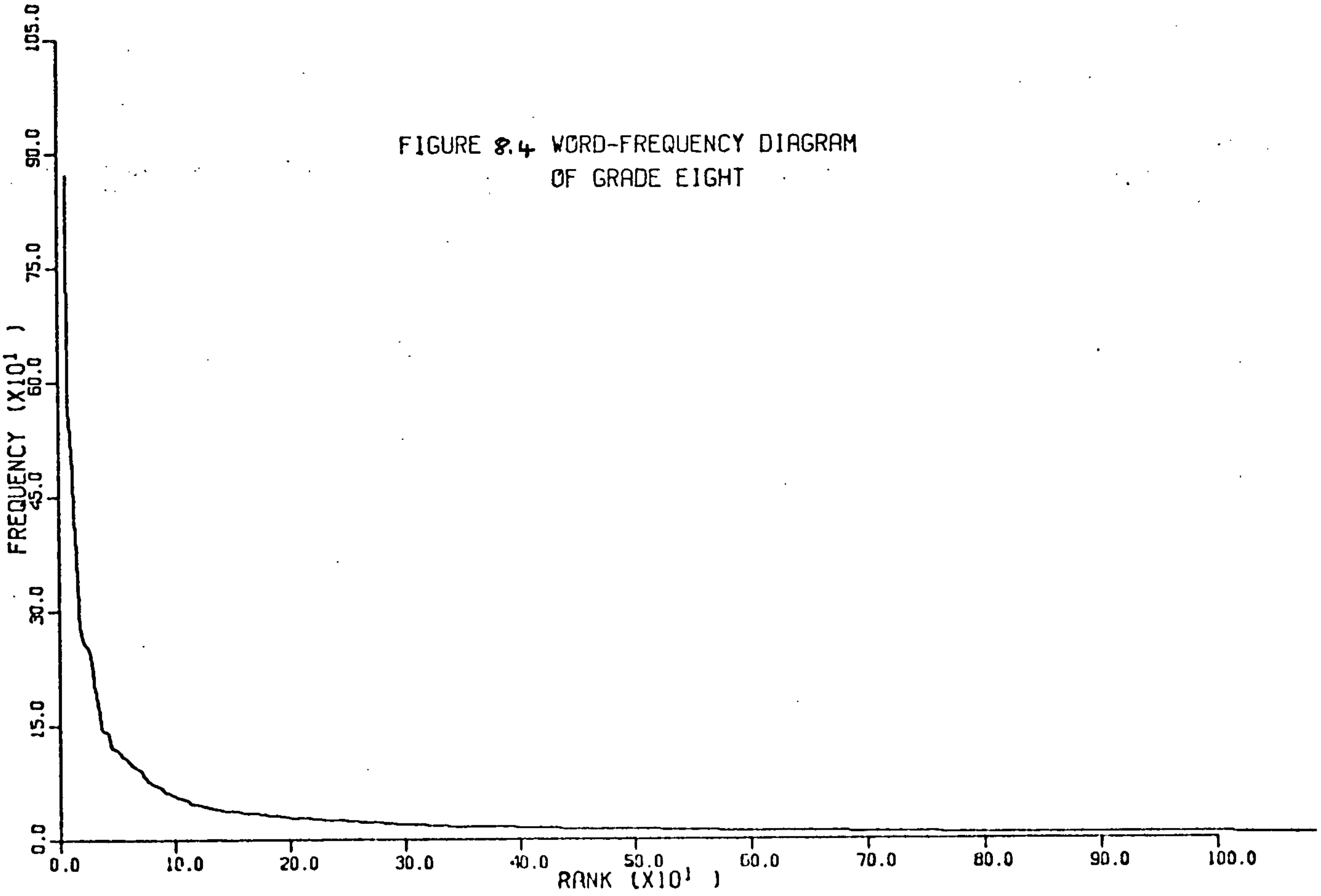


FIGURE 8.4 WORD-FREQUENCY DIAGRAM
OF GRADE EIGHT
FREQUENCY (X10¹)
RANK (X10¹)
0.0
15.0
30.0
45.0
60.0
75.0
90.0
105.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0

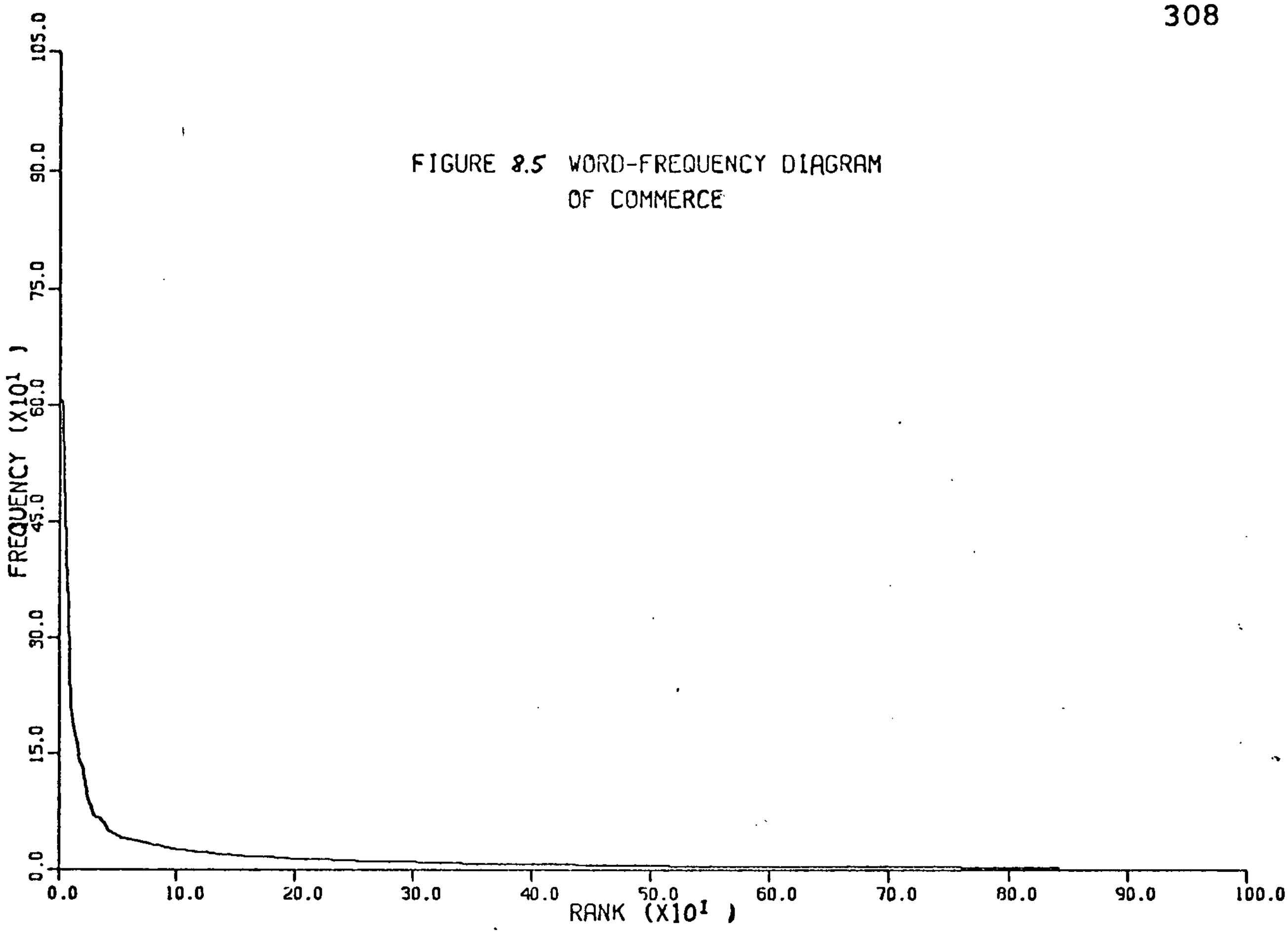


FIGURE 8.5 WORD-FREQUENCY DIAGRAM
OF COMMERCE
FREQUENCY (X10¹)
RANK (X10¹)
0.0
15.0
30.0
45.0
60.0
75.0
90.0
105.0
10.0
20.0
40.0
50.0
70.0
80.0
100.0

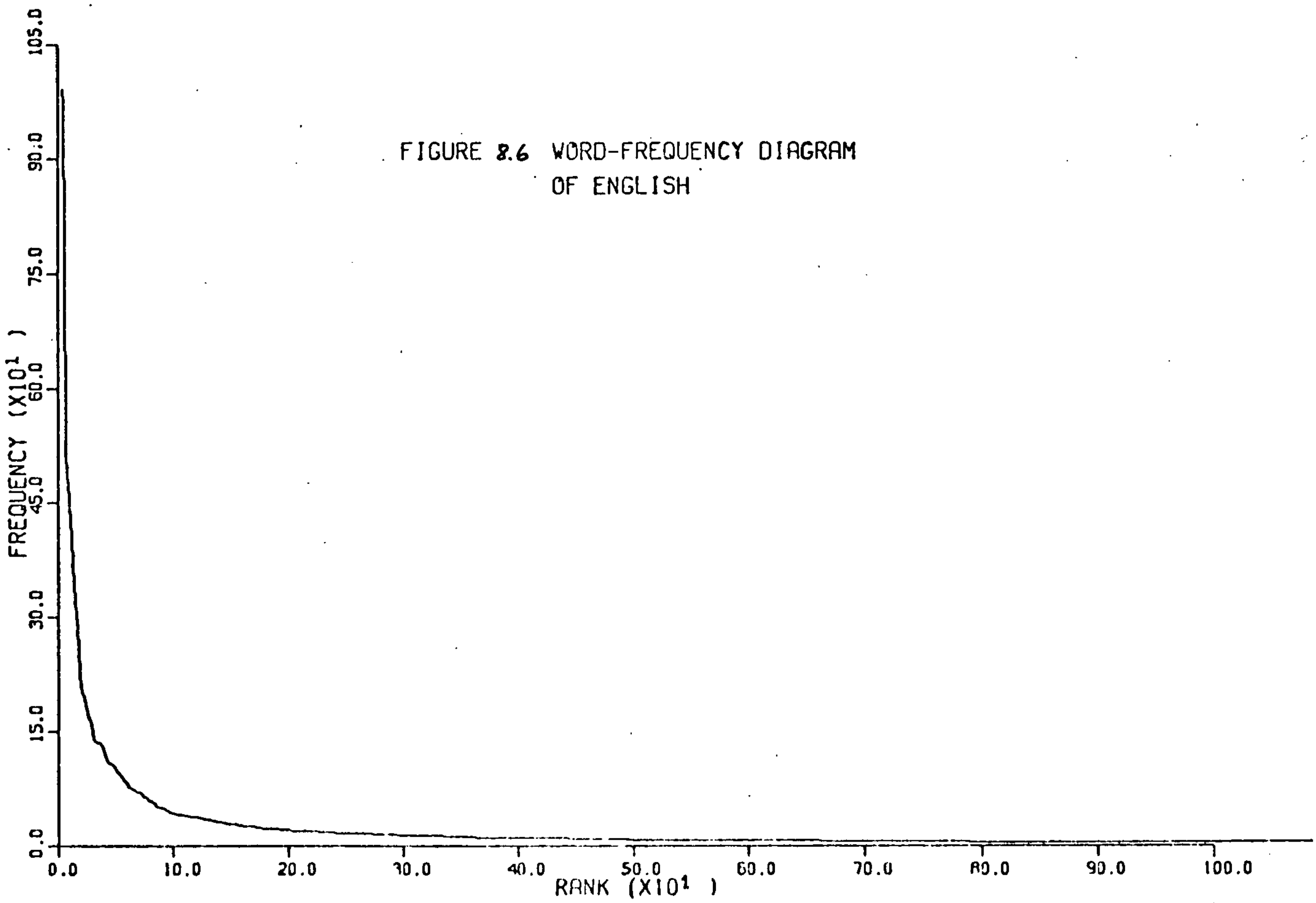


FIGURE 8.6 WORD-FREQUENCY DIAGRAM
OF ENGLISH
FREQUENCY (X10¹)
RANK (X10¹)
0.0
15.0
30.0
45.0
60.0
75.0
90.0
105.0
10.0
20.0
40.0
50.0
70.0
80.0
100.0

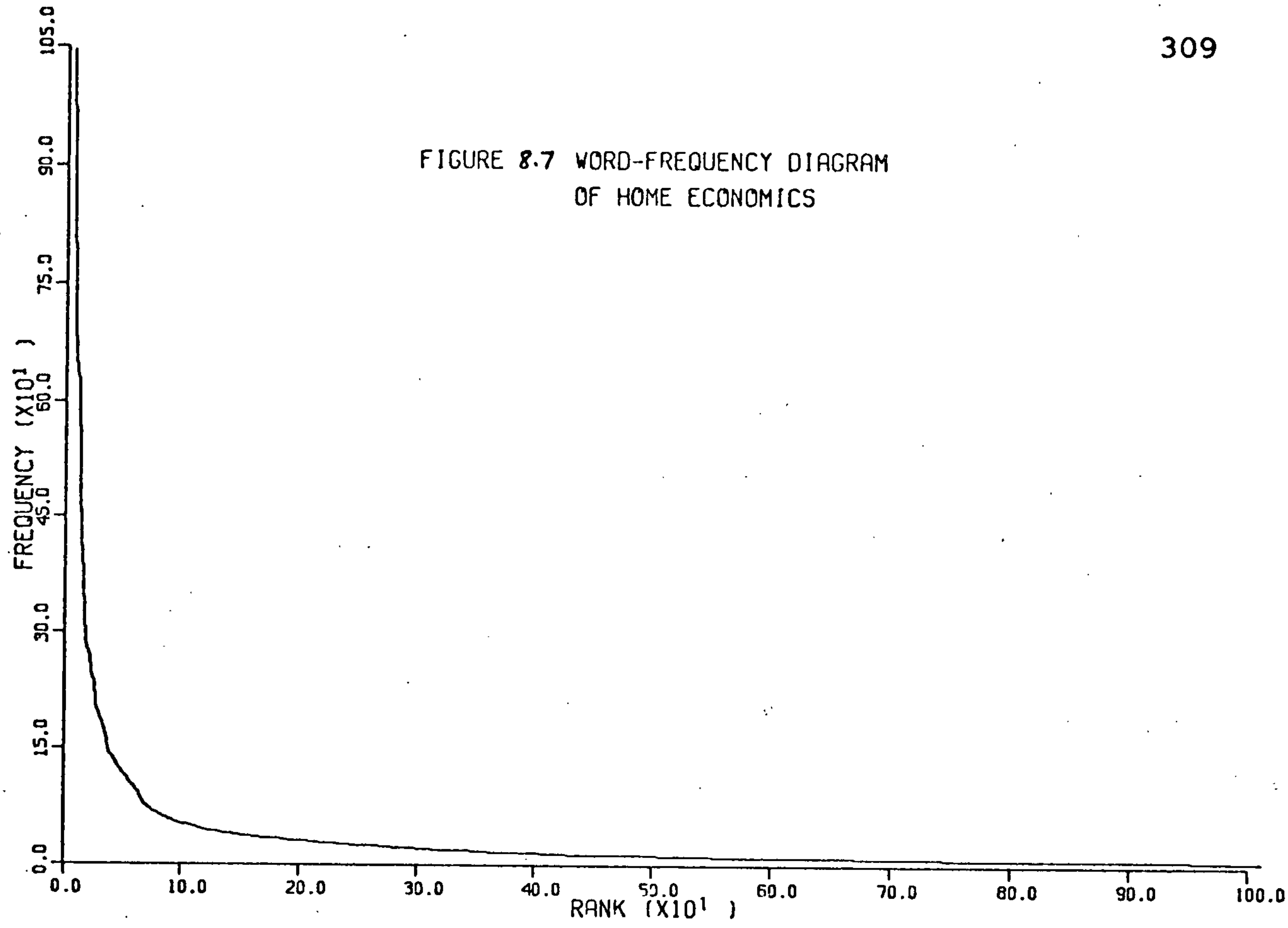


FIGURE 8.7 WORD-FREQUENCY DIAGRAM OF HOME ECONOMICS

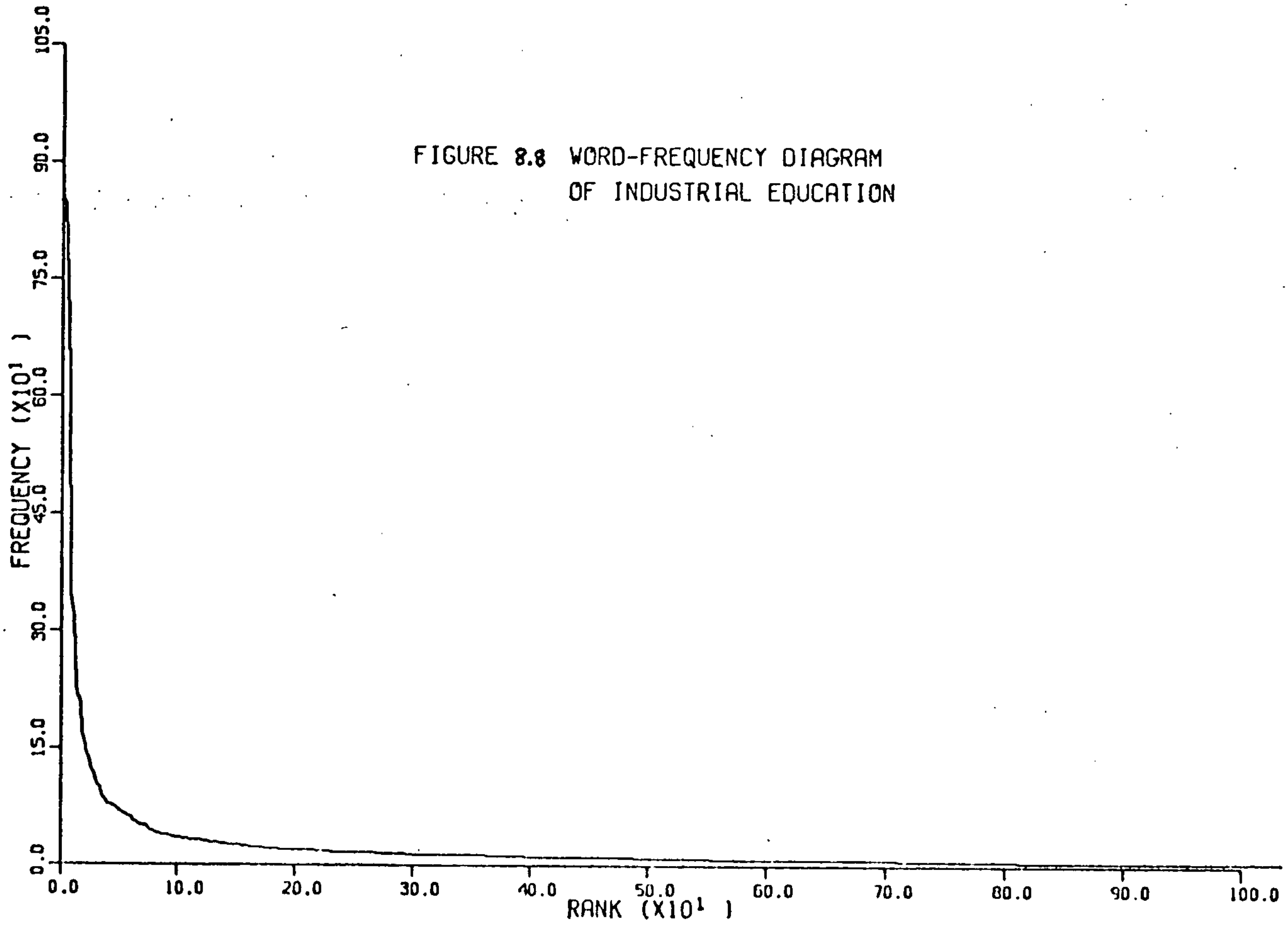


FIGURE 8.8 WORD-FREQUENCY DIAGRAM OF INDUSTRIAL EDUCATION

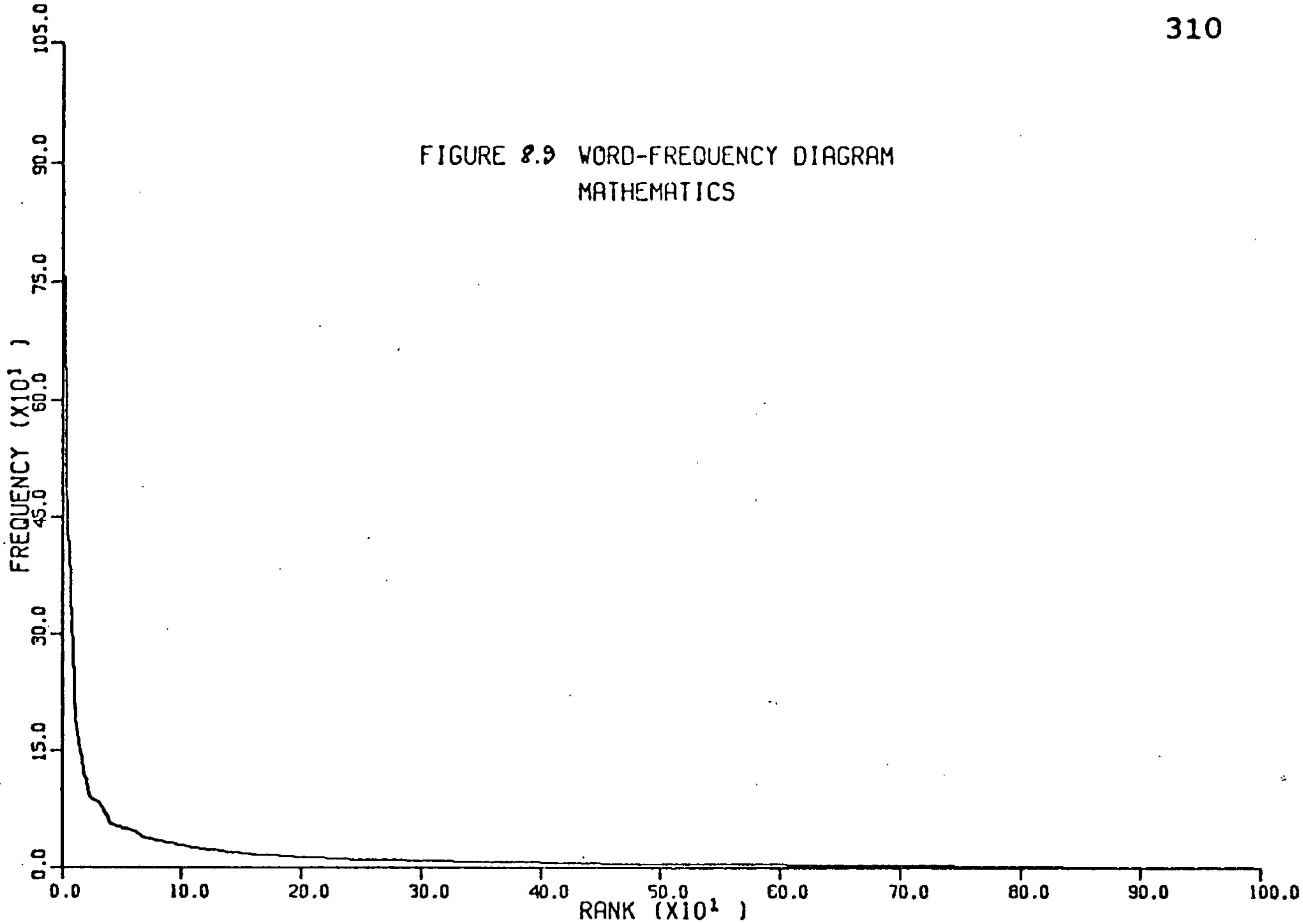


FIGURE 8.9 WORD-FREQUENCY DIAGRAM
MATHEMATICS
FREQUENCY (X10^1)
RANK (X10^1)
0.0
15.0
30.0
45.0
60.0
75.0
90.0
105.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0

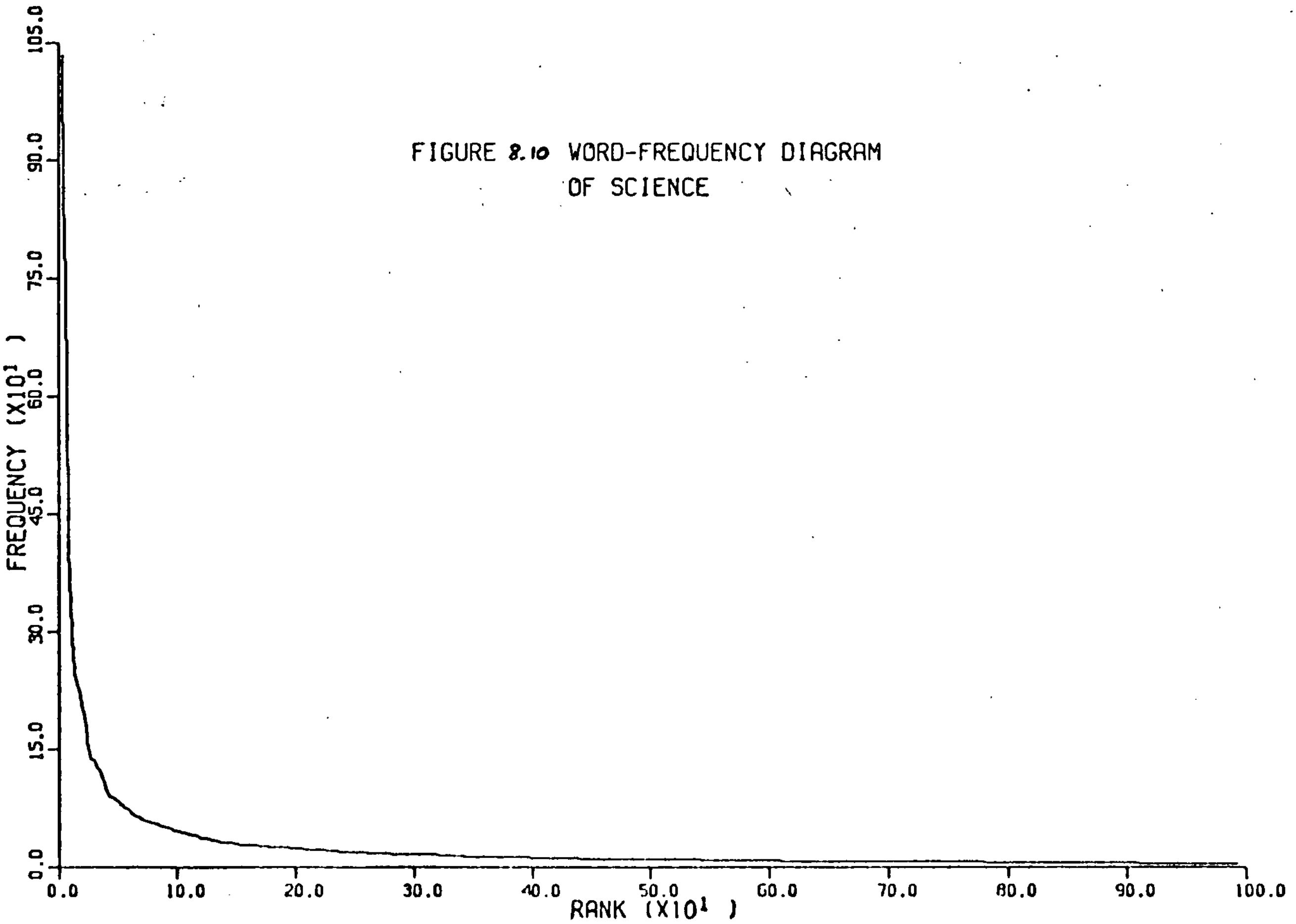


FIGURE 8.10 WORD-FREQUENCY DIAGRAM
OF SCIENCE
FREQUENCY (X10^1)
RANK (X10^1)
0.0
15.0
30.0
45.0
60.0
75.0
90.0
105.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0

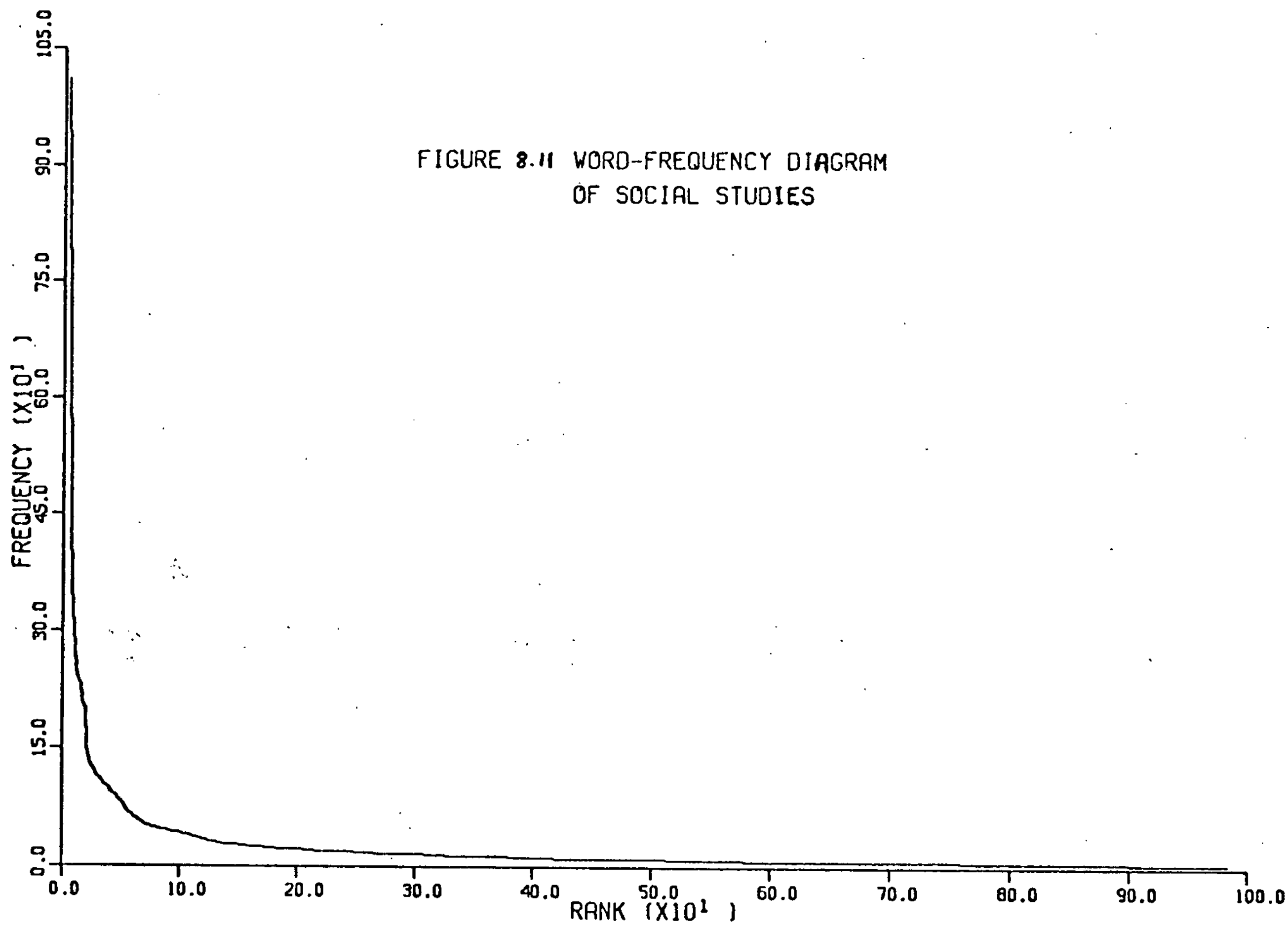


FIGURE 8.11 WORD-FREQUENCY DIAGRAM
OF SOCIAL STUDIES
FREQUENCY (X10¹)
RANK (X10¹)
0.0
15.0
30.0
45.0
60.0
75.0
90.0
105.0
10.0
20.0
40.0
50.0
70.0
80.0
100.0

PROFESSIONAL WRITING

A. ARTICLES

"Aren't We Being Conned?" The B.C. Teacher, 48, (May-June, 1969), 327-328.

"Patterns in Literature" B.C. English Teacher, 10, (June, 1970), 102-106.

"PANORAMA - A Study Technique" Journal of Reading, 17, (November, 1973), 132-135.

"Some Problems in Secondary Reading" Fourth Annual Reading Conference, U.B.C., Vancouver, Centre for Continuing Education, 1973.

"The Effect of Idioms on Children's Reading and Understanding of Prose" Teachers, Tangibles, Techniques: Comprehension of Content in Reading, ed. Bonny Schulwitz (Newark: International Reading Association, 1973).

"Idioms and Reading Comprehension" Journal of Reading Behavior, 10, (Fall, 1974), 30-36.

B. REFERENCE TEXTS

Summers, Edward G., Brother Leonard Courtney., and Peter Edwards. Guide to Professional Textbooks and Research in Secondary Reading, University of British Columbia, Information Research Centre, 1973.

C. RESEARCH REPORTS

A Study of the Effectiveness of the Vancouver School Board Reading Centre Program. Research Report 73-10, Vancouver, Board of School Trustees, June 1973.

An Interaction-Network Instrument for Measuring Pupil-Interaction in a Learning Environment. Research Report 73-15, Vancouver, Board of School Trustees, July 1973.

An Evaluation of the Communications English Program at Britannia Secondary School. Research Report 73-16, Vancouver, Board of School Trustees, July 1973.

An Evaluation of the English 11E Program at Templeton Secondary School. Research Report 73-19, Vancouver, Board of School Trustees, July 1973.

An Evaluation of the Adaptability of Grade 8 Students to the University Hill Secondary Program. Research Report 73-22, Vancouver, Board of School Trustees, July 1973.