TM-1908/200/00

Stylistic Analysis

Report on the Second Year of Research

1 March 1966

TM-1908/200/00

TECHNICAL MEMORANDUM

(TM Series)

The research reported herein was conducted under SDC's independent research program and Contract Nonr-4427(00), Office of Naval Research, U. S. Navy, Task No. NR 348-005.

Stylistic Analysis

Report on the Second Year of Research

by

Sally Yeates Sedelow

March 1, 1966

SYSTEM DEVELOPMENT CORPORATION
2500 COLORADO AVE
SANTA MONICA
CALIFORNIA
90406

(page 2 blank)

Stylistic Analysis

Report on the Second Year of Research

by

Sally Yeates Sedelow

ABSTRACT

This document describes the current capabilities of the computer-based programs being developed for the analysis of style. The power of the programs for content analysis, semantic analysis and stylistic discrimination is evaluated on the basis of their analytical runs on natural language text. The value of such output for political and diplomatic negotiations, as well as for more long-range projects involving information retrieval and automatic language production, etc., is illustrated.

PREFACE

This is the second annual report on research in stylistic analysis being carried out at the System Development Corporation under a grant from the Office of Naval Research, Information Systems Branch, Contract Nonr-4427(00). I wish to thank the Office of Naval Research for its generous financial assistance; I am also grateful to the staff members of the Information Systems Branch for the helpful suggestions and advice they have provided during the course of this Project. Thanks are also due to the System Development Corporation for administrative and financial aid as well as for many other kinds of research support; space prohibits listing the names of the many people at SDC who have in some way contributed to this effort. Acknowledgement should also be made to Saint Louis University for freeing me from some teaching responsibilities so as to provide time for research, and to a number of friends and colleagues in the wider academic community who have made helpful comments and suggestions on research results and procedures. Primary among the latter is my husband, Walter A. Sedelow, Jr., who shares my interest in the analysis of language and has conducted parallel research in this area.[1]

Very special thanks are due Terry Ruggles of the System Development Corporation who has done all the program coding and has contributed many innovations of his own to the program design. In addition, he has looked

---

[1] Walter A. Sedelow, Jr., "Science and the Language of History," Behavioral Science, 2(1), January 1957, pp 80-82; History as Language, An Essay at a New Historiography..., Harvard University Archives, Widener Library, Cambridge, Massachusetts, 1957.

after numerous on-the-scene details associated with the project, which have included the producing of listings, flowcharts, etc., for documents, for reports, and for distribution to individuals doing research in the general area of linguistic analysis.

(page 6 blank)

TABLE OF CONTENTS

I. Introduction

Although the philosophy underlying this research has been described in some detail elsewhere[2] a brief resume seems desirable here in order to provide an orientation for this report.

We have defined style, in this context, as the patterns formed in the linguistic encoding of information; stylistic analysis, then, is the perception of pattern in language. As it is transmitted, information inevitably assumes some shape or pattern. The information itself, as well as whatever syntactic, rhythmic, etc., properties it takes on, helps provide the shape or pattern. For example, Ellegard's work[3] on the authorship of the Junius letters was based on various authors' choices of information-bearing words (as opposed to words at least primarily fulfilling some syntactic or dictional function within the sentence). Ellegard discovered that he could discriminate one author from another on the basis of style, or pattern, of information-word choice. Also in support of the contention that information is a part of, rather than apart from, style, it is worth noting that discussions of form (a term which has been frequently used in

---

[2] Sally Y. Sedelow and Walter A. Sedelow, Jr., "A Preface to Computational Stylistics," in J. Leed (Ed), The Computer and Literary Style: Eight Introductory Essays and Studies, Kent State University Press, 1966 (also available as SDC document SP-1534, 17 February 1964); S. Y. Sedelow, W. A. Sedelow, Jr., and T. L. Ruggles, "Some Parameters for Computational Stylistics," Proceedings, IBM Literary Data Processing Conference, IBM, Yorktown Heights, September, 1964.

[3] Alvar Ellegard, A Statistical Method for Determining Authorship: The Junius Letters, 1769-1772, Goteborg, 1962.

talk of style) often focus upon shifts in the presentation of an idea: if a writer "reverses" himself through some type of contrast or antithesis, sometimes it is, in fact, the perception of such a shift in the information-bearing words that enables the linguistic analyst to talk about the shape or form of the idea. Thus, this research project is based on the premise that the choice of information-bearing words, as well as all the other patterns which they help form as they are embedded in sentences, paragraphs, speeches, chapters, etc., is the necessary province of stylistic analysis. This approach is conformal with those comparatively recent developments in stylistics which imply, or require, the termination of the very long-standing style-content or form-content dichotomies as invalid analytically and/or heuristically. Strictly speaking, these divisions appear untenable; style and content are not only not mutually exclusive but, heuristically speaking, look as though they are only effectively used when mutually inclusive.

The study of style has many different applications for the information analyst. It is important for content analysis and information retrieval. The study of patterns of information-bearing (content) words reveals with a potential degree of rigor hitherto unknown, the important, organizing concepts in a given body of text. The perceeption of these patterns is essential for the researcher attempting to make a precise assessment of the import of a given document. Information retrieval, based on an evaluation of a document's relevance for a particular search request, is also dependent on the exact recognition of concepts or clusters of concepts.

The study of style is the key to the discrimination of one author from another. That an author's style is his signature has already been demonstrated in the case of Swift,[4] Madison and Hamilton,[5] and Sir Philip Francis.[6] In each of these instances, a different procedure for discriminating style was used: Milic concentrated on syntax, Mosteller and Wallace on the use of certain function words, and (as already mentioned) Ellegard upon content words. This variety of techniques resulting in successful discrimination suggests the need for a computer-oriented analytical "package" which can quickly and efficiently take as many measures as needed to effect identification of individual styles.[7] The importance of such identification to an information analyst is immense: for example, shifts in authorship within a given document or report may imply interpolations representing a change in policy or policy sensitivities. The quick perception of such shifts or sensitivities may well provide valuable leads as to the appropriate directions for diplomatic or military initiative.

---

[4]Louis T. Milic, A Quantitative Approach to the Style of Jonathan Swift, (Microfilm), University Microfilm, Ann Arbor, Michigan, 1964.

[5]Frederick Mosteller and David L. Wallace, Inference and Disputed Authorship: The Federalist, Addison-Wesley Publishing Co., Inc., Reading, Massachusetts, 1964.

[6]Ellegard, op.cit.

[7]Robert Wachal, of the University of Iowa, is working on a related problem. He is interested in establishing the minimum set of measures for the successful allocation of all disputed literary paternity cases.

The study of style is also of major importance to the automatic production of language. The much-predicted direct verbal communication between man and computer will be successful only if the computer's verbal styles are sufficiently varied and idiomatic to permit quick and accurate comprehension. Adequate machine translations and automatic abstracts will be achieved only if attention is paid to the style which is idiomatic to the group for whom the translation or abstract is intended. Superior machine translations will strive for the optimum "mix" between the style of the original and the style of the language to which the intended reader is accustomed; and, for various purposes, we may wish to produce near-simultaneous variant translations--each adjusted to the verbal behavior patterns of a particular population. It is, to understate, not too early to begin studies of style directed toward these goals.

The bulk of this document describes results of research undertaken thus far, some programming procedures used for the research, and directions the research will take during the coming year. The discussion of VIA, one of the two program systems used for this research, will not include detailed computer program descriptions; these are available in SDC publications TM-1908/100/00 and TM-1908/009/00 (both documents are necessary for adequate understanding of the program and its revisions). The computer program for MAPTEXT, the second of the systems, has not been described elsewhere and will be included in this document. All descriptions of computer programs are the joint effort of Sally Sedelow and Terry Ruggles. A summary of professional activities during the past year follows the program description of MAPTEXT.

II. Research Procedures and Results

The stylistic analysis research project has entailed the development of two programs, VIA and MAPTEXT. VIA, which is an acronym for Verbally-Indexed Associations, searches a given document for conceptually associated word patterns; MAPTEXT provides both graphic and statistical descriptions of the document, using as part of its input the output from the VIA program.

A. VIA

As will be evident from examples later in this section, VIA groups words in lists in a manner analogous to groupings in conventional thesauri such as Roget's. That is, all the words in a given main list have a reasonably close semantic relationship to each other and words on any given list or lists attached to the main list have a similarly close semantic relationship among themselves as well as a semantic relationship (sometimes very close, sometimes tangential) with the words on the main list. As VIA was originally conceived, all words in a given body of text were grouped together by root and further searches of the text for possible-associated-words were based on these root groups. In an effort to prevent the output from being too dense[8] and some of the material retrieved from being extraneous, all searches were word-specific.[9] It became apparent, however, that although

[8] By "density," we mean the number of lists and of words on those lists. We feared that an extremely large number of these lists and words would interfere with ready perception of verbal patterns.

[9] "Word-specific" means the identical form, rather than root form.

this procedure guarded against the hazards cited above, it also resulted in the loss of relevant information. Therefore, during the course of this year the THESAUR section of the VIA program has been revised to provide for more comprehensive retrieval of primary words (words are designated as "primary" by the researcher on the basis of frequency, usually either high or low) and associated words (words in lists linked to the primary words).[10] Conceptualized abstractly, the change might be described as follows:

Old Procedure

P1

A1

P2(A1)

A2

P3(A2)

A3

P1 is the primary word. A1 is the list of textual words which are associated with P1. P2 is a member of A1 and is also a primary word in its own right. A2 is a list of textual words which are associated with P2, etc.

[10] For a complete description of program revisions, see S. Y. Sedelow and T. Ruggles, "Updating of THESAUR Program," TM-1908/009/00, System Development Corporation, Santa Monica, California, 17 December 1965.

New Procedure

P1 (R1)

A1 (R1...n)

P2 (A1) (R2)

(A2) (R1...n)

P3 (A2) (R3)

A3 (R1...n)

The (R1), (R2), and (R3) following the entries for (respectively) P1, P2, and P3 indicate an exploration of the input text for cognates of the primary word. If, for example, the computer's thesaurus contains the word "man" as a primary word and the section of text currently being examined does not have "man" but does contain "manly," the latter would be recognized as having the same root as "man," and a search of the text would be instituted for all words associated with "man" in the computer's thesaurus. The (R1...n) following the entries for A1, A2, and A3 indicates an exploration of the text for root group members for each of the occurring associated words. This addition to the THESAUR program guards against the loss of thematic, or ideational, continuities owing to shifts in form in any of the primary or associated words involved in the search.

A number of changes have been made in VIA's output format. These changes are described in the document mentioned earlier (TM-1908/009/00), and they will be apparent in examples of the output shown in this section.

VIA's output can be discussed with reference to at least three aspects: 1) content analysis, 2) semantic analysis, 3) stylistic discrimination. These aspects, not always mutually exclusive, will be discussed in sequence, and followed by a description of plans for further research.

1. Content Analysis. One of VIA's purposes is to pick out important ideas or concepts within a given body of text. As has already been implied, the output is based on the designation (as primary) of certain root groups in the text. In order to make this designation, the researcher may request a computer run which produces a description of the entire text like that shown in the example in Figure 1.

```
8 ACCOMPLISHED      (0001--0058--0001--0060)(0001--0060--0002--0043)(0001--0068--0002--0014)
                    (0001--0068--0003--0014)(0001--0075--0001--0009)(0001--0083--0003--0025)
                    (0001--0130--0003--0012)(0001--0187--0002--0010)(0001--0235--0002--0012)

  ACCOMPLISHMENT    (0001--0025--0002--0031)(0001--0123--0001--0033)

  ACCOMPLISHMENTS   (0001--0027--0002--0008)(0001--0042--0001--0055)(0001--0150--0001--0028)
                    (0001--0286--0001--0006)

  ACCOMPLISH        (0001--0074--0003--0025)(0001--0229--0001--0046)(0001--0252--0003--0026)
                                                                                        15

9 ACCORDANCE        (0001--0030--0001--0002)(0001--0190--0003--0068)

  ACCORDING         (0001--0020--0001--0022)(0001--0020--0001--0030)(0001--0071--0001--0050)
                    (0001--0073--0001--0025)(0001--0073--0001--0032)(0001--0166--0003--0032)
                    (0001--0166--0004--0030)(0001--0173--0001--0008)(0001--0190--0003--0056)
                    (0001--0234--0004--0024)(0001--0254--0001--0001)(0001--0263--0002--0006)

  ACCORD            (0001--0105--0001--0032)
                                                                                        13
```

Figure 1. VIA Output Showing
Root Groups, Indices, and Frequencies

The numbers within parentheses show where in the text the word has occurred (for example, the first occurrence of ACCOMPLISHED is in chapter 1, paragraph 58, sentence 1, word 60). The arabic number to the left of ACCOMPLISHED is the MAT(Match)CNT(Count) of its root group.[11] Other words with a MATCNT of 8 are ACCOMPLISHMENT, ACCOMPLISHMENTS, and ACCOMPLISH. ACCORDANCE begins a new root group with a MATCNT of 9. MATCNTs are used by the computer program for sorting. Their value in this output is simply to provide visual separation of the root groups. The arabic number below the final index entry for ACCOMPLISH indicates how many times the words in this root group occurred in the portion of text indexed. Thus, there are 15 occurrences of words in the ACCOMPLISHED root group, and 13 occurrences of words in the ACCORDANCE root group. Function words are usually excluded from this output, although they may be included if the researcher so chooses.

Given the above output, the researcher may then decide which words he wishes to designate as primary. Since human perception of important concepts within a text is often a function of sufficient repetition to make an impression on the reader, words occurring with high frequency relative to the rest of the text are usually designated as primary. Thus for the first

---

[11] In the SUFFIX section of the VIA system, all words with the same root are assigned an identical MATCNT number, indicating that they have matching roots.

chapter in both the RAND translation[12] and the Praeger translation[13] of Sokolovsky's book on Soviet military strategy, all words occurring 50 or more times were specified as primary words. On other, often unspecified, bases these words would be cited as the key to the major, or organizing, ideas in that chapter. Obviously, the researcher can alter the density of his output simply by raising or lowering the cut-off point determining the designation of primary words.

Having determined which words are to be called primary, the researcher then manually consults synonym dictionaries, thesauri, and (using the index) the contexts in which the word occurs, in order to compile a list of words which, if they occur in the text, should be associated with the primary word. The computer then institutes a rather complicated search for these possible-associated-words, searching first for each word associated with the given primary word, next for words linked to such of those associated words as have primary status in their own right, etc.[14] This linking procedure now goes down through five levels; the computer compiles a history of the links which subsequently can be used in runs on the next section of text. Thus, if a linking word does

---

[12] V. D. Sokolovsky, Soviet Military Strategy, Dinerstein, Goure, and Wolfe (Trans.), Prentice-Hall, Inc., 1963.

[13] V. D. Sokolovsky, Military Strategy: Soviet Doctrine and Concepts, Translation Services Branch, Foreign Technology Division, Wright-Patterson AFB (Trans.), Frederick A. Praeger, Inc., 1963. Because the title of the translating service is so long, the phrase "Praeger translation" is used throughout this report when referring to this translation of Sokolovsky.

[14] The results of this linking procedure will be clear in the examples appearing in Figure 2 et passim of this report.

not occur in the next section of text but words associated with that link do, the latter words are retrieved and linked back to the appropriate primary word; in this way, the continuity of a concept or idea, if there is continuity, is shown. The initial thesaurus is also saved, so that the computer may consult its own thesaurus during subsequent runs on additional segments of the text. Thus, through the accretion of information, both through its own linking procedures (the human researcher provides only the first-level links) and the storing of thesaurus input, the computer increasingly takes over the manual part of the program. Nonetheless, the output of the program looks so promising that it would seem worthwhile to begin the effort to automate that first, currently manual, search as well. For a further discussion of this problem, see pages 71-72 of this report.

Having in mind this general description of the way VIA works, we may now turn to the output, examining it as a form of content analysis. Two texts have been used for experimental purposes. One text comprises the two translations, already mentioned, of Sokolovsky, and the other is the Kittredge edition of Hamlet (Ginn and Company). The latter was selected as a "control" because of the vast amount of work in the public domain on this particular play of Shakespeare's, and because of its presumed stylistic distance from the translations of Sokolovsky.

In order to test the hypothesis that the perception of organizing concepts (or, in the case of Hamlet and other literature, some might say "themes") is based upon the repetition of words and their cognates, no effort was made to achieve a dense output for either of the texts. The

cut-off point used to designate primary words in Hamlet--10 or more occurrences--might seem low when compared with the 50 or more occurrences used for Sokolovsky. But in all of Hamlet, which comprises approximately 17,104 content word occurrences, only 66 words occur 10 or more times (this number does not include proper names). As one might expect, the literary artist varies his word choice more than the writer on military strategy. Indeed, if--as will be apparent later in this report is indeed the case--clues as to the major themes in a literary work can be gained from this approach, the case for VIA's efficacy with other kinds of text would, a fortiori, seem easily made.

A detailed analysis of all of VIA's output from Hamlet and the implications of this output for the interpretation of the play would take much space. Nonetheless, it does seem desirable to reproduce some of the output in order to support comments about the operation of VIA. For that purpose, all words in Act V which are either primary words in the act or have qualified as primary words in earlier acts will be shown along with their linked lists of associated words.[15] The output is shown in Figure 2.

[15] Conventional act designations have been followed for this experimental research even though Shakespeare did not provide these designations. For purposes of discussion and criticism, however, the tradition of referring to material within given acts is well established.

```
 19 *ACTION                                                        6
                      DEED
                      MIRROR
              --------PERFORM
    *                 PLAY                             7
                              SPEECH
              --------RANT
              --------STAGE
     ACT
     ACTS
 24 *AGAIN                                                         5
 29  ALL'S                                                        12
     ALL
 73 *AY                                                            7
149 *CANNOT                                                        5
187  COME                                                         32
    *                 GO                               6
                              COME
                              FAREWELL
    *                         LEAVE                    3
                                      COME
                                      FAREWELL
                                      GO
              ------------------------QUIT
                                      STOP
     COMES
     COMING
242 *DEAR                                                          3
243 *DEATH                                                         8
              --------BLEED
              --------BORN
              --------BURIAL
                      DEAD
```

Figure 2. VIA Output from Hamlet, Act V.

Number at left is MATCNT. Numbers at right are frequency counts for occurrences of words in primary root group. Note that some associated words are also primary words. Primary words preceded by asterisks belong to root groups which occur fewer than 10 times in Act V; however they have occurred 10 or more times in an earlier act or acts. Words preceded by dashes have not occurred in any of the preceding acts. All lists linked to primary root groups should be thought of as linked to all words in the group. Thus, in the first group shown, DEED, MIRROR, PERFORM, etc., are linked to ACTION, ACT and ACTS.

```
                    DIED
                    DIE
                    GRAVE
                    LIFE
            --------MURD'ROUS
                    MURTHER
            --------OBSEQUIES
                    PLOT
            --------ROT
                    SLAIN
     DEATHS
265  DID                                                    13
287  DO                                                     22
                    DID
                    DIDST
            --------DOEST
                    DOST
     *              DOTH                          2
                            DO
            --------UNDO
290  *DOTH                                                  2
                    DO                            22
                            DID
                            DIDST
            ----------------DOEST
                            DOST
                            DOTH
            ----------------UNDO
298   DRINK                                                 11
            --------CUP
            --------LIQUOR
            --------POTION
      DRINKS
312  *EARTH                                                 9
     *              HEAVEN                                 5
                            EARTH
                            STARS
                    LAND
     *              NATURE                                 5
                            LIFE
                            UNNATURAL
                    WORLD
```

Figure 2. (Continued)

```
327  *ENDS                                                        1
                  BEGIN
                  EAR
                  LAST
                  STOP
353  *EYE                                                         1
                  BODY
                  SEE                                    10
                           EYE
                           SIGHT
                  ---------VIEW
356  *FAIR                                                        4
366  *FATHER'S                                                    2
                  MAN                                    11
                           FATHER
                           FELLOW
                           GENTLEMAN
                           WOMAN
     *            MOTHER                                  3
                           FATHER
                  ---------FEMALE
                           SON
                           WOMAN
                  SON
399  *FOLLOW                                                      6
                  COMPLY
                  OBEY
422  *FRIENDS                                                     1
                  ENEMY
446   GIVE                                                       12
448  *GOD                                                         2
                  DEVIL
451   GOODMAN                                                    20
                  EVIL
                  NOBLE
      GOOD
453  *GO                                                          6
                  COME                                   32
                           GO
                  FAREWELL
     *            LEAVE                                   3
                           COME                           32
                                    GO
                           FAREWELL
                           GO
                  ---------QUIT
                           STOP
```

Figure 2. (Continued)

```
476   HAMLET'S                                                    26
      HAMLET
485   HATH                                                        14
491   *HEARERS                                                     7
                    EAR
                    HEARD
              ------LEARN
                    UNDERSTAND
      HEARING
      HEAR
492   *HEART                                                       4
                    BODY
                    FEELING
                    LIFE
      *             SOUL                          4
                          BODY
                          CONSCIENCE
                          HEART
494   *HEAVEN                                                      5
      *             EARTH                         9
                          HEAVEN
                          LAND
      *                   NATURE                  5
                                LIFE
                                UNNATURAL
                          WORLD
                    STARS
      HEAVENS
498   HERE'S                                                      27
                    PRESENT
      HERE
512   HORATIO                                                     11
577   KING'S                                                      18
                    MAJESTY
      *             QUEEN                         7
                          KING
                          MAJESTY
              ----------ROYAL
                          WOMAN
                    ROYAL
      KINGDOM
      KING
      KINGS
```

Figure 2. (Continued)

```
585   KNOWING                                                       19
                          IGNORANCE
              --------LEARNING
                          UNDERSTAND
      KNOWN
      KNOW
      KNOWS
589   LAERTES                                                       13
603  *LEAVE                                                          3
                          COME                            32
     *                             GO                      6
                                             COME
                                             FAREWELL
                                             LEAVE
                          FAREWELL
     *                    GO                               6
                                   COME                   32
                                             GO
                                   FAREWELL
                                   LEAVE
              --------QUIT
                          STOP
      LEAVES
608   LET                                                           23
                          VOUCHSAFE
611   LIE                                                           12
      LIES
      LIEST
627  *LOOK                                                           2
                          CARRIAGE
                          COMPLEXION
                          SEE                             10
     *                             EYE                     1
                                             BODY
                                             SEE
                                   SIGHT
              ----------------VIEW
631  *LOSE                                                           2
                          GAIN
      LOSING
635  *LOVE                                                           8
                          DESIRE
              --------DOTE
                          HOT
                          PASSION
```

Figure 2. (Continued)

```
636   MADNESS                                                    11
                  ILL
      MAD
641   MAN'S                                                      11
      *           FATHER                          2
                        MAN
      *                 MOTHER                    3
                                  FATHER
                ------------------------FEMALE
                                  SON
                                  WOMAN
                        SON
                  FELLOW
                  GENTLEMAN
                  WOMAN
      MAN
652   *MEAN                                                      3
                --------INSTRUMENT
                  PURPOSE
      MEANT
679   MOST                                                       10
                  MANY
680   *MOTHER'S                                                  3
      *           FATHER                          2
                        MAN                       11
                                  FATHER
                                  FELLOW
                                  GENTLEMAN
                                  WOMAN
                        MOTHER
                        SON
                --------FEMALE
                  SON
                  WOMAN
      MOTHER
694   *NATURE                                                    5
                  LIFE
                  UNNATURAL
702   *NIGHT'S                                                   3
                --------DARK
                  DAY
                  FEAR
                  LIGHT
      NIGHT
```

Figure 2. (Continued)

709	NO				30
		NAY			
710	*NOTHING				4
		--------EMPTY			
744	*OUT				9
	OUTWARD				
788	*PLAY				7
		SPEECH			
812	*PRAYERS				4
		BEG			
		BESEECH			
	*	GOD		2	
			DEVIL		
		--------PRIEST			
	PRAY				
841	*QUEEN				7
		KING		18	
			MAJESTY		
			QUEEN		
			ROYAL		
		MAJESTY			
		--------ROYAL			
		WOMAN			
963	SAY				11
		SAY'ST			
	*	SPEAK		9	
			MOUTH		
			SAY		
			SPEECH		
			TONGUE		
			VOICE		
		TELL		15	
		----------------	REPORT		
		----------------	SAY		
	SAYS				
977	SEEN				10
	*	EYE		1	
			BODY		
			SEE		
		SIGHT			
		--------VIEW			
	SEE				

Figure 2. (Continued)

```
1003  *SHOWING                                                      3
      *                PLAY                                 7
                                SPEECH
       SHOW
       SHOWS
1023  *SLEEP                                                        1
               --------BORN
      *                DEATH                                8
               ----------------BLEED
               ----------------BORN
               ----------------BURIAL
                                DEAD
                                DIED
                                DIE
                                GRAVE
                                LIFE
               ----------------MURD'ROUS
                                MURTHER
               ----------------OBSEQUIES
                                PLOT
               ----------------ROT
                                SLAIN
                       DIE
                       LIVE
                       SLAIN
1036  *SOUL                                                         4
                       BODY
                       CONSCIENCE
      *                HEART                                4
                                BODY
                                FEELING
                                LIFE
                                SOUL
       SOULS
1041  *SPEAK                                                        9
                       MOUTH
                       SAY                                 11
                                SAY'ST
                                SPEAK
                                TELL                       15
               ------------------------REPORT
               ------------------------SAY
                       SPEECH
                       TONGUE
                       VOICE
       SPEAKS
```

Figure 2. (Continued)

1082	*SWEET					7
	*	DEAR			3	
	SWEETS					
1091	TELL					15
		--------REPORT				
		--------SAY			11	
			SAY'ST			
	*		SPEAK		9	
				MOUTH		
				SAY		
				SPEECH		
				TONGUE		
				VOICE		
			TELL			
1105	*THINK					6
		BELIEVE				
		REASON				
	*	THOUGHT			2	
		----------------	IMPRESSION			
1112	*THOUGHT					2
		--------IMPRESSION				
1123	TIME					11
		TERM				
	TIMES					
1127	*TOO					8
1142	*TRUE					2
		--------ERROR				
		--------EXACT				
		FAITHFUL				
		--------VERITY				
	TRUTH					
1191	*WATCH					1
1200	*WELCOME					1
1201	WELL					16
		--------FLOURISH				
		HEALTH				
		ILL				
1235	*WORDS					1
		SAY			11	
			SAY'ST			
	*		SPEAK		9	
				MOUTH		
				SAY		
				SPEECH		

Figure 2. (Continued)

```
                                 TONGUE
                                 VOICE
                         TELL                     15
         ------------------------REPORT
         ------------------------SAY
*                SPEAK                            9
                         MOUTH
                         SAY                     11
                                 SAY'ST
                                 SPEAK
                                 TELL            15
         --------------------------------REPORT
         --------------------------------SAY
                         SPEECH
                         TONGUE
                         VOICE
                 TERM
         --------WRITE
```

Figure 2. (Continued)

Among the root groups which qualify as primary in Act V (all those root groups at the left for which no asterisk is associated with any word(s) in the group), a few are present only because VIA does not currently take care of the problem presented by form shifts accompanying tense shift or because the function word list used for this experimental run does not take care of archaic forms. Thus DID and HATH both appear as primary words although neither would seem to be of primary interest to the researcher.

A few other preliminary comments concerning the output should be made. First, the lists upon which the output is based contain antonyms as well as synonyms because both are important to the analysis of any given concept. Second, although proper names are listed in the output, no sublists are linked to them simply because it was felt that the contents of any such list would be completely dependent upon the imagination of the researcher and, therefore, totally subjective; anyone who wanted such lists in the output could, of course, provide for them. (One way to make such lists a trifle less subjective, of course, would be to use as possible-associated-words only those words which occur within the sentence or speech in which the name appears. But, even so, some selection would doubtless be necessary.) Third, a number of lists contain information that seems repetitive so far as content analysis is concerned (especially good examples are the entries under COME and GO); the repetition was tolerated, however, because of its possible value for semantic analysis (see the discussion on pages 58-62 of this report). Fourth, in a few cases the presence of a given word on a list may seem surprising, if not unwarranted. An example is the occurrence of

EAR on a list headed by ENDS. EAR qualifies for the list through its meaning as a fruit, or end, of a particular cycle--in this case, the growing cycle. This linking is derived from a conventional thesaurus, and not from semi-free-associating on the part of the researcher.

Assuming that the researcher knows nothing about the play (and this is the basis on which VIA might well be used for many purposes of information analysis), a glance at the primary words for Act V would suggest, for example, that a drink or the act of drinking was of some importance in this act. The presence of the word POTION on the list linked to DRINK might imply something unusual about the drink. The researcher would guess that a KING had an important role in this act and he would note that the words associated with KING have connotations of a king's court and of his kingdom. Among proper names used in this act, he would find that the names of Hamlet, Horatio, and Laertes qualify as primary words, with Hamlet having the highest frequency. The researcher would notice that the more generic word MAN occurs with some frequency and that some of the words associated with it imply familial relationships. If he were to glance at those words in the output which had qualified as primary words in earlier acts, the researcher would discover some of the familial terms: FATHER, DAUGHTER, and MOTHER. He would take note of the fact that these familial terms no longer had primacy in the final act (in fact, DAUGHTER does not occur at all) but that the more general, more impersonal terms KING and MAN did. If the word POTION associated with DRINK had overtones of the unusual, the presence of MADNESS

as a primary word must amplify those tones. The word ILL linked to MADNESS may suggest disease, as does the linking of ILL to WELL.[16]

Another primary word of possible interest in this act is the word TELL. The researcher would notice that a number of words associated with TELL suggest an emphasis upon speaking or talking. He would note, in fact, that the word SAY also qualifies as a primary word in this act and that the word SPEAK has had primacy in an earlier act or acts, as has the word WORDS. There would seem to be an interest in language, especially the spoken language (the word WRITE appears for the first time, and only once, in this final act), as well as perhaps an interest in reporting or relaying something. The latter interest seems especially important in this last act, with the emphasis upon TELL (a glance at the output from earlier acts would show the researcher that TELL had not qualified as a primary word in those acts). The researcher would also notice a premium upon knowledge: the words KNOW, KNOWS, KNOWN, and KNOWING all occur as well as the words LEARNING and UNDERSTAND which are associated with them. The contrasting word IGNORANCE is also present. Although the word DEATH does not quite qualify as a primary

---

[16] WELL is an example of a word which qualifies as a primary word because it is multiple-faceted as to function and meaning. In Hamlet, WELL is used, among other ways, as an interjection (as in "Well, again"), as an idiom implying some sort of assent or agreement ("Very well"), and as a descriptor of health or condition. Thus, if this primary word group appeared in isolation, its implications as to content would have to be investigated thoroughly by examination of context. However, because of the presence of MADNESS as a primary word, and the high incidence of words associated with DEATH in this chapter, the presence of ILL as an antonym to WELL would not seem too misleading so far as a general assessment of content (or "theme" or "tone") was concerned.

word in this act (it occurs eight times), 14 of the words on the computer's thesaurus list which might be associated with death occur in this act. An examination of this list reveals words which definitely imply violence: BLEED, MURD'ROUS, MURTHER, and SLAIN. The word PLOT, which also appears on this list, may provide some insight into the importance of knowledge, of KNOWING. In keeping with the theme of violence, it may be of interest that three of the words associated with LOVE, which does not quite qualify as a primary word, imply violence rather than tenderness. The fourth word, DOTE, also suggests a lack of restraint although not of the same sort as that connoted by DESIRE, HOT, and PASSION. Other words among those which do not qualify as primary words in this act but did qualify earlier, which catch the eye because of the relatively large number of words linked to them, include SLEEP, SOUL, HEAVEN, EARTH, HEARERS, and ACT. In this act, SLEEP occurs just once, but even so, there is nothing of restfulness or soothingness in the words associated with it. Rather, there is the violence associated with DEATH which is linked to SLEEP. The presence of words clustered around SOUL, HEAVEN, and EARTH indicates that Hamlet is not simply a description of derring-do for the sake of derring-do. The words FEELING, CONSCIENCE, and HEART are associated with SOUL. Words associated with both HEAVEN and EARTH include NATURE, WORLD, STARS, LIFE, and UNNATURAL. When, bearing these groups in mind, the researcher sees that GOD and THINK and THOUGHT have also qualified as primary words in earlier acts, and that in this act the words BELIEVE and REASON are linked to THINK, he might suspect that there is some emphasis upon abstract speculation and upon speculation

about the abstract in this play. Considering the already noted emphasis upon saying and speaking, the presence of HEARERS as a sometime primary word will not surprise the researcher. He might also suspect that ACT, another earlier primary word, has some connection with SPEAK; the word STAGE, which is linked to ACT would strengthen this suspicion.

High-frequency words which seem unequivocally to be verbs--words such as COME, DO, and GO--would present more difficulty. The researcher might conclude that there is a good deal of coming and going but he would not be able to attach these actions to individuals. If he knew anything about Shakespeare, he would realize that many of the verbal references to coming and going are simply a part of Shakespeare's staging technique. These verbs provide stage directions in the text itself and, as well, they help to fill intervals during which characters are making entrances and exits. Lacking this information, the researcher would need to turn to the context of the occurrences of the verbs in order to determine who was coming and going and the reasons for the activity.

Even after this quick initial glance at VIA's output from Act V, the researcher would have a good notion of important content, both as to plot and theme, of the fifth act of Hamlet. This is the act in which the King (Claudius) successfully plots to kill Hamlet. The murder will occur during a duel, which Hamlet supposes to be a friendly sporting match, between Hamlet and Laertes. The tip of Laertes' foil will have been coated with a poison which, once the skin has been penetrated, will bring almost immediate death. Although Laertes is an excellent fencer, the King takes the

additional precaution of preparing a poison potion (had the researcher's suspicions been aroused by POTION he could have consulted the complete output for Act V and discovered that both POISON'D and POISON occur) should Hamlet have occasion to drink a toast in victory. In the general carnage at the end of the play, Hamlet, Laertes, and the King are all dispatched by the poisoned foil and the Queen, in an effort to drink to Hamlet's early success in the duel, dies of the poisoned potion; she is ignorant of the plot against Hamlet. In fact, the King is the only character who knows all the details of the plot as well as of his own earlier criminal act (the murder of Hamlet's father); the incomplete knowledge of the other characters permits the tragic denouement of Hamlet. The final act's emphasis upon TELL is clearly pointed when Hamlet, dying, urges Horatio to "Absent thee from felicity awhile, /And in this harsh world draw thy breath in pain, /To tell my story."

Details of the thematic implications of the primary word clusters in Act V would entail too much space and time for this report. Suffice it to say, that placed within the context of VIA's output from other acts, these primary word clusters would lead the researcher to most, if not all, of the themes literary critics have found important in Hamlet; in addition, some appear, such as the familial and non-familial designations of characters briefly mentioned on page 30 of this report, upon which no critics seem to have commented but which, nonetheless, would considerably enrich interpretations of the play.

For an even greater enrichment of analytical power (and this is important for the information analyst who is seeking a more subtle interpretation or examination of a work with which he is already acquainted), the output could be made more dense either by lowering the cut-off point for designation of primary words or by using an entire cluster (such as that associated with DEATH) as a point for further departure.[17] Even without going this far, a much more detailed picture of Hamlet than that hinted at in the initial glance above could be achieved on the basis of the output already obtained. The researcher would first want to examine the output of Act V more carefully; he would particularly want to compare VIA's work on this act with that on the other acts in order to get a sense of differences and similarities. Thus, he could place his observations on Act V within the context of the entire play and trace the development of themes or the varying emphases upon those themes. The comparison of outputs from the various sections of a play, or document, etc., could be automated so as to make the similarities and differences stand out sharply; we plan to add this capacity to the VIA system.

This rather detailed discussion of VIA's performance on a small part of Hamlet was undertaken to show what VIA can do with a text representing a highly skilled, varied style; on this rather stiff test, VIA's performance

---

[17] That is, because so many words associated with DEATH occur, this concept should qualify for further exploration even though DEATH, itself, occurs only eight times and is, therefore, not primary.

seems impressive. Now let us look at VIA's output for a text closer to the kind of data with which information analysts must more often cope. Reproduced in Figure 3 is a partial output from the Praeger translation of the first chapter in Soviet Military Strategy.

```
15  *ACID                                                          1
              --------SHARP
              --------TEST
20  *ACUTE                                                         1
              --------POINTED
              --------SHARP
44   AIMED                                                        53
     AIM
              --------ASPIRE  (DIFF. FORM APPEARS IN TEXT)
              --------BEARING
              --------COURSE
              --------DESIGN
              --------DIRECT
              --------END
              --------GOAL
              --------INTEND  (DIFF. FORM APPEARS IN TEXT)
              --------OBJECTIVE
              --------OBJECT
              --------POINT
              --------PURPOSE
              --------TARGET
     AIMS
55   *ALL                                                          42
              --------ALTOGETHER
              --------COMPLETE
              --------ENTIRE
              --------EVERY
              --------INTEGRATE  (DIFF. FORM APPEARS IN TEXT)
              --------NO
              --------TOTALITY
              --------UNITY                                        78
                      --------ACCORD
                      --------AGREEMENT
                      --------DISAGREE  (DIFF. FORM APPEARS IN TEXT)
                      --------SOLIDARITY
                      --------UNDERSTANDING
              --------WHOLE
```

Figure 3. Partial VIA Output for Sokolovsky, Praeger Translation

For an explanation of notation in this output which differs from that in Hamlet, see commentary immediately following this Figure.

```
80  *APPARENTLY                                                          3
          --------APPEARS
          --------CLEARLY
          --------EVIDENTLY  (DIFF. FORM APPEARS IN TEXT)
    APPARENT
83  *APPRECIATED                                                         1
          --------ESTIMATED  (DIFF. FORM APPEARS IN TEXT)
          --------REALIZED
          --------UNDERSTAND  (DIFF. FORM APPEARS IN TEXT)
93   ARMAMENT                                                          144
     ARMAMENTS
     ARMED
     ARMIES
     ARMS
     ARMY'S
     ARMY
    (DIFFERENT FORM APPEARS IN THESAURUS)
          --------DISARM  (DIFF. FORM APPEARS IN TEXT)
          --------FIREARM
          --------GUN
          --------MISSILE
          --------NUCLEAR
          --------ROCKET
          --------STEEL
          --------WAR                                                  293
                  --------AVIATION
                  --------BALLISTICS
                  --------BATTLEFIELD
                  --------BATTLE
                  --------BELLIGERENT
                  --------CAMPAIGN
                  --------COMBAT
                  --------CONFLICT                                      24
                          --------BATTLE
                          --------DISAGREEMENT
                          --------OPPOSE  (DIFF. FORM APPEARS IN TEXT)
                          --------WAR
                  --------FIGHTING  (DIFF. FORM APPEARS IN TEXT)
                  --------GUNPOWDER
                  --------HOSTILITIES
                  --------MILITARY                                     285
                          --------FIGHTER (DIFF. FORM APPEARS IN TEXT)
                          --------GENERAL                              53
                                  --------COMMON
                                  --------NATURAL
                                  --------NORMAL (DIFF. FORM APPEARS IN TEXT)
                                  --------REGULAR (DIFF. FORM APPEARS IN TEXT)
                                  --------TOTAL
                                  --------TYPICAL
                                  --------UNIVERSAL
                                  --------WHOLE
```

Figure 3. (Continued)

```
                                --------MAJOR
                                --------MARSHAL
                                --------OFFICER
                                --------SOLDIER
                                --------WARFARE
                                --------WAR
                        --------PACIFIC
                        --------PEACE
                        --------STRATEGY                                    283
                                --------GENERALSHIP (DIFF. FORM APPEARS IN TEXT) 53
                                        --------COMMON
                                        --------NATURAL
                                        --------NORMAL (DIFF. FORM APPEARS IN TEXT)
                                        --------REGULAR (DIFF. FORM APPEARS IN TEXT)
                                        --------TOTAL
                                        --------TYPICAL
                                        --------UNIVERSAL
                                        --------WHOLE
                                --------PLAN
                                --------TACTICS
                        --------TACTICS
                --------WEAPON
102  *ASCRIBED                                                              1
                --------ASSIGNED
                --------ATTRIBUTED (DIFF. FORM APPEARS IN TEXT)
144  *BEFOREHAND                                                            1
                --------EARLY
                --------LATE
                --------PRIOR
172  *BOOST                                                                 1
                --------IMPETUS
                --------RAISE
185  *BRILLIANT                                                             1
                --------CREATIVE (DIFF. FORM APPEARS IN TEXT)
                --------CREATIVITY
                --------INTELLIGENT (DIFF. FORM APPEARS IN TEXT)
220  *CESSATION                                                             1
                --------CONTINUATION
                --------STOPPAGE (DIFF. FORM APPEARS IN TEXT)
                --------STOP
235  *CHRONOLOGY                                                            1
                --------SEQUENCE
                --------TIMING
252  *CODIFIED                                                              1
255  *COINCIDE                                                              1
                --------SIMULTANEOUS
256  *COLLECTION                                                            1
                --------GATHERING (DIFF. FORM APPEARS IN TEXT)
                --------SUMMARY (DIFF. FORM APPEARS IN TEXT)
288   CONFLICTING                                                           24
     *CONFLICT
                --------BATTLE
```

Figure 3. (Continued)

```
        --------DISAGREEMENT
        --------OPPOSE (DIFF. FORM APPEARS IN TEXT)
        --------WAR                                                    293
                --------AVIATION
                --------BALLISTICS
                --------BATTLEFIELD
                --------BATTLE
                --------BELLIGERENT
                --------CAMPAIGN
                --------COMBAT
                --------CONFLICT
                --------FIGHTING (DIFF. FORM APPEARS IN TEXT)
                --------GUNPOWDER
                --------HOSTILITIES
                --------MILITARY                                       285
                        --------FIGHTER (DIFF. FORM APPEARS IN TEXT)
                        --------GENERAL                               53
                                --------COMMON
                                --------NATURAL
                                --------NORMAL (DIFF. FORM APPEARS IN TEXT)
                                --------REGULAR (DIFF. FORM APPEARS IN TEXT)
                                --------TOTAL
                                --------TYPICAL
                                --------UNIVERSAL
                                --------WHOLE
                        --------MAJOR
                        --------MARSHAL
                        --------OFFICER
                        --------SOLDIER
                        --------WARFARE
                        --------WAR
                --------PACIFIC
                --------PEACE
                --------STRATEGY                                      283
                        --------GENERALSHIP (DIFF. FORM APPEARS IN TEXT) 53
                                --------COMMON
                                --------NATURAL
                                --------NORMAL (DIFF. FORM APPEARS IN TEXT)
                                --------REGULAR (DIFF. FORM APPEARS IN TEXT)
                                --------TOTAL
                                --------TYPICAL
                                --------UNIVERSAL
                                --------WHOLE
                        --------PLAN
                        --------TACTICS
                --------TACTICS
     CONFLICT
304  *CONTEMPORARY                                                     2
        --------CURRENT
```

Figure 3. (Continued)

```
308  *CONTINGENCIES                                                    1
              --------CIRCUMSTANCES
              --------EVENTS
              --------INCIDENTS (DIFF. FORM APPEARS IN TEXT)
              --------OCCURRENCES (DIFF. FORM APPEARS IN TEXT)
315  *CONVENIENCE                                                      1
321  *CORE                                                             1
              --------BASIS
              --------CENTER
328   COUNTRIES                                                      122
      COUNTRY'S
      COUNTRY
              --------COLONY (DIFF. FORM APPEARS IN TEXT)
              --------CONTROL
              --------DICTATION
              --------DOMINATION
              --------INFLUENCE
              --------MASTERY
              --------PATRIOTIC
              --------POLITY (DIFF. FORM APPEARS IN TEXT)
              --------POWER
              --------REALM
              --------SOVEREIGNTY
              --------STATE                                            95
                      --------CONDITION
                      --------COUNTRY
                      --------GOVERNMENT
                      --------REALM
334  *CRIPPLING                                                        1
337  *CRYSTALLIZATION                                                  1
              --------FORMING
340  *CUSTOMARILY                                                      1
              --------NORMALLY (DIFF. FORM APPEARS IN TEXT)
              --------REGULARLY
              --------USUALLY
351  *DEBATE                                                           1
              --------AGREE (DIFF. FORM APPEARS IN TEXT)
              --------ARGUE (DIFF. FORM APPEARS IN TEXT)
              --------DISCUSS
              --------DISPUTE
355  *DECREASE                                                         1
              --------ADD (DIFF. FORM APPEARS IN TEXT)
              --------DEVELOP                                          73
                      --------ADVANCE
                      --------EVOLUTION (DIFF. FORM APPEARS IN TEXT)
                      --------EVOLVE (DIFF. FORM APPEARS IN TEXT)
                      --------EXPAND (DIFF. FORM APPEARS IN TEXT)
                      --------MATURE (DIFF. FORM APPEARS IN TEXT)
                      --------PROGRESS
              --------GROW (DIFF. FORM APPEARS IN TEXT)
              --------INCREASE
366  *DEMISE                                                           1
```

Figure 3. (Continued)

```
374  *DEPARTURE                                                          2
              --------REFERENCE (DIFF. FORM APPEARS IN TEXT)
              --------START
389   DEVELOPED                                                         73
      DEVELOPING
      DEVELOPMENT
      DEVELOPMENTS
      DEVELOP
              --------ADVANCE
              --------EVOLUTION (DIFF. FORM APPEARS IN TEXT)
              --------EVOLVE (DIFF. FORM APPEARS IN TEXT)
              --------EXPAND (DIFF. FORM APPEARS IN TEXT)
              --------MATURE (DIFF. FORM APPEARS IN TEXT)
              --------PROGRESS
      DEVELOPS
406  *DISCREPANCIES                                                      1
              --------CONFORM
              --------CONTRADICTION
              --------DIVERSITIES (DIFF. FORM APPEARS IN TEXT)
408  *DISINTEGRATING                                                     1
              --------ARM                                          147
                      --------DISARM (DIFF. FORM APPEARS IN TEXT)
                      --------FIREARM
                      --------GUN
                      --------MISSILE
                      --------NUCLEAR
                      --------ROCKET
                      --------STEEL
                      --------WAR                                  293
                              --------AVIATION
                              --------BALLISTICS
                              --------BATTLEFIELD
                              --------BATTLE
                              --------BELLIGERENT
                              --------CAMPAIGN
                              --------COMBAT
                              --------CONFLICT                      24
                                      --------BATTLE
                                      --------DISAGREEMENT
                                      --------OPPOSE (DIFF. FORM APPEARS IN TEXT)
                                      --------WAR
                              --------FIGHTING (DIFF. FORM APPEARS IN TEXT)
                              --------GUNPOWDER
                              --------HOSTILITIES
                              --------MILITARY                     285
                                      --------FIGHTER (DIFF. FORM APPEARS IN TEXT)
                                      --------GENERAL               53
                                              --------COMMON
                                              --------NATURAL
                                              --------NORMAL (DIFF. FORM APPEARS IN TEXT)
```

Figure 3. (Continued)

```
                                        --------REGULAR
                                                (DIFF. FORM APPEARS IN TEXT)
                                        --------TOTAL
                                        --------TYPICAL
                                        --------UNIVERSAL
                                        --------WHOLE
                                --------MAJOR
                                --------MARSHAL
                                --------OFFICER
                                --------SOLDIER
                                --------WARFARE
                                --------WAR
                        --------PACIFIC
                        --------PEACE
                        --------STRATEGY                                  283
                                --------GENERALSHIP                        53
                                        (DIFF. FORM APPEARS IN TEXT)
                                        --------COMMON
                                        --------NATURAL
                                        --------NORMAL
                                                (DIFF. FORM APPEARS IN TEXT)
                                        --------REGULAR
                                                (DIFF. FORM APPEARS IN TEXT)
                                        --------TOTAL
                                        --------TYPICAL
                                        --------UNIVERSAL
                                        --------WHOLE
                                --------PLAN
                                --------TACTICS
                        --------TACTICS
                --------WEAPON
    *       --------CRIPPLE (DIFF. FORM APPEARS IN TEXT)                    1
            --------DEMORALIZING
            --------ENABLE
            --------INCAPACITATE
            --------STRENGTHEN
            --------UNDERMINE
409 *DISORDER                                                             1
    *       --------DISORGANIZED                                          1
                --------DISORDER
                --------METHOD
                --------ORDER
                --------ORGANIZED (DIFF. FORM APPEARS IN TEXT)
                --------SYSTEM
            --------METHOD
            --------ORDER
            --------ORGANIZE
            --------SYSTEM
410 *DISORGANIZED                                                         1
    *       --------DISORDER                                             1
                --------DISORGANIZED
```

Figure 3. (Continued)

```
                    --------METHOD
                    --------ORDER
                    --------ORGANIZE
                    --------SYSTEM
            --------METHOD
            --------ORDER
            --------ORGANIZED (DIFF. FORM APPEARS IN TEXT)
            --------SYSTEM
411 *DISPERSAL                                                    3
            --------CALL
            --------CITE
            --------DIVISION
            --------GATHER (DIFF. FORM APPEARS IN TEXT)
            --------MUSTER (DIFF. FORM APPEARS IN TEXT)
            --------SEPARATE
    DISPERSED
    DISPERSING
412 *DISPLACEMENT                                                 1
            --------ESTABLISH (DIFF. FORM APPEARS IN TEXT)
            --------OUTMODE (DIFF. FORM APPEARS IN TEXT)
            --------PLACE
            --------REPLACEMENT
            --------SITUATE (DIFF. FORM APPEARS IN TEXT)
413 *DISPLAYED                                                    1
            --------EXPOSE (DIFF. FORM APPEARS IN TEXT)
            --------HIDE
            --------MANIFEST
            --------MASK
            --------SHOWN
420  DIVERSIONARY                                                 2
    *DIVERT
            --------TURN
426 *DOLLARS                                                      3
            --------FINANCE
            --------MONEY
431 *DOUBT                                                        1
            --------CERTAINTY (DIFF. FORM APPEARS IN TEXT)
            --------INDECISIVE
434 *DRAFT                                                        1
435 *DRASTICALLY                                                  1
            --------CONSIDERABLY
            --------GREATLY
438 *DROP                                                         1
            --------INCREASE
            --------RAISE
            --------RISE
439 *DRUG                                                         1
            --------INTOXICATE
            --------PACIFY (DIFF. FORM APPEARS IN TEXT)
450  ECONOMIC                                                   130
            --------MONEY
            --------TRADE
```

Figure 3. (Continued)

```
      ECONOMICS
      ECONOMIES
      ECONOMIST
      ECONOMY
601   FORCED                                                          105
      FORCE
              --------CONTROL
              --------MIGHT
              --------POWERFUL (DIFF. FORM APPEARS IN TEXT)
              --------POWERLESS (DIFF. FORM APPEARS IN TEXT)
              --------POWER
              --------STRENGTH
              --------WEAK
      FORCES
      FORCIBLY
648   GENERALIZATION                                                   53
      GENERALIZATIONS
      GENERALIZED
      GENERALIZE
      GENERALIZING
      GENERAL
              --------COMMON
              --------NATURAL
              --------NORMAL (DIFF. FORM APPEARS IN TEXT)
              --------REGULAR (DIFF. FORM APPEARS IN TEXT)
              --------TOTAL
              --------TYPICAL
              --------UNIVERSAL
              --------WHOLE
      GENERALS
931   MILITARILY                                                      285
      MILITARISTIC
      MILITARIST
      MILITARISTS
      MILITARY
              --------FIGHTER (DIFF. FORM APPEARS IN TEXT)
              --------GENERAL                                          53
                      --------COMMON
                      --------NATURAL
                      --------NORMAL (DIFF. FORM APPEARS IN TEXT)
                      --------REGULAR (DIFF. FORM APPEARS IN TEXT)
                      --------TOTAL
                      --------TYPICAL
                      --------UNIVERSAL
                      --------WHOLE
              --------MAJOR
              --------MARSHAL
              --------OFFICER
              --------SOLDIER
              --------WARFARE                                         293
                      --------AVIATION
                      --------BALLISTICS
```

Figure 3. (Continued)

```
                        --------BATTLEFIELD
                        --------BATTLE
                        --------BELLIGERENT
                        --------CAMPAIGN
                        --------COMBAT
                        --------CONFLICT                                    24
                                --------BATTLE
                                --------DISAGREEMENT
                                --------OPPOSE (DIFF. FORM APPEARS IN TEXT)
                                --------WAR
                        --------FIGHTING (DIFF. FORM APPEARS IN TEXT)
                        --------GUNPOWDER
                        --------HOSTILITIES
                        --------MILITARY
                        --------PACIFIC
                        --------PEACE
                        --------STRATEGY                                   283
                                --------GENERALSHIP (DIFF. FORM APPEARS IN TEXT) 53
                                        --------COMMON
                                        --------NATURAL
                                        --------NORMAL
                                                (DIFF. FORM APPEARS IN TEXT)
                                        --------REGULAR
                                                (DIFF. FORM APPEARS IN TEXT)
                                        --------TOTAL
                                        --------TYPICAL
                                        --------UNIVERSAL
                                        --------WHOLE
                                --------PLAN
                                --------TACTICS
                        --------TACTICS
                --------WAR
 952    MORALE                                                             63
                --------ENTHUSIASM
                --------HIGH
                --------LOW
                --------SPIRIT
1046    OPERATIONAL                                                        62
        OPERATION
                --------FLANKING (DIFF. FORM APPEARS IN TEXT)
                --------MANEUVER
                --------MILITARY                                          285
                        --------FIGHTER (DIFF. FORM APPEARS IN TEXT)
                        --------GENERAL                                   53
                                --------COMMON
                                --------NATURAL
                                --------NORMAL (DIFF. FORM APPEARS IN TEXT)
                                --------REGULAR (DIFF. FORM APPEARS IN TEXT)
                                --------TOTAL
                                --------TYPICAL
                                --------UNIVERSAL
```

Figure 3. (Continued)

```
                --------WHOLE
        --------MAJOR
        --------MARSHAL
        --------OFFICER
        --------SOLDIER
        --------WARFARE                                          293
                --------AVIATION
                --------BALLISTICS
                --------BATTLEFIELD
                --------BATTLE
                --------BELLIGERENT
                --------CAMPAIGN
                --------COMBAT
                --------CONFLICT                                  24
                        --------BATTLE
                        --------DISAGREEMENT
                        --------OPPOSE (DIFF. FORM APPEARS IN TEXT)
                        --------WAR
                --------FIGHTING (DIFF. FORM APPEARS IN TEXT)
                --------GUNPOWDER
                --------HOSTILITIES
                --------MILITARY
                --------PACIFIC
                --------PEACE
                --------STRATEGY                                283
                        --------GENERALSHIP                      53
                                (DIFF. FORM APPEARS IN TEXT)
                                --------COMMON
                                --------NATURAL
                                --------NORMAL
                                        (DIFF. FORM APPEARS IN TEXT)
                                --------REGULAR
                                        (DIFF. FORM APPEARS IN TEXT)
                                --------TOTAL
                                --------TYPICAL
                                --------UNIVERSAL
                                --------WHOLE
                        --------PLAN
                        --------TACTICS
                --------TACTICS
        --------WAR
OPERATIONS
OPERATIVE
1129    POLICY'S                                                 159
POLICY
        --------DESIGN
        --------OUTLINE
        --------PLAN
        --------PROGRAM
POLITICALLY
POLITICAL
POLITICIANS
```

Figure 3. (Continued)

```
        POLITICS
1393    SOCIALISM                                                          60
                --------COMMUNIST
                --------COOPERATIVE (DIFF. FORM APPEARS IN TEXT)
                --------PARTY
        SOCIALIST
        SOCIALIZATION
        SOCIALIZED
        SOCIAL
1414    SOVIET                                                             54
                --------COMMUNISM
                --------RUSSIA
                --------SOCIALIST                                          60
                        --------COMMUNIST
                        --------COOPERATIVE (DIFF. FORM APPEARS IN TEXT)
                        --------PARTY
1432    STATE'S                                                            95
        STATED
        STATEMENT
        STATE
                --------CONDITION
                --------COUNTRY                                           122
                        --------COLONY (DIFF. FORM APPEARS IN TEXT)
                        --------CONTROL
                        --------DICTATION
                        --------DOMINATION
                        --------INFLUENCE
                        --------MASTERY
                        --------PATRIOTIC
                        --------POLITY (DIFF. FORM APPEARS IN TEXT)
                        --------POWER
                        --------REALM
                        --------SOVEREIGNTY
                        --------STATE
                --------GOVERNMENT
                --------REALM
        STATES'
        STATESMEN
        STATES
1443    STRATEGICALLY                                                     283
        STRATEGIC
        STRATEGISTS
        STRATEGOS
        STRATEGY
                --------GENERALSHIP (DIFF. FORM APPEARS IN TEXT)          53
                        --------COMMON
                        --------NATURAL
                        --------NORMAL (DIFF. FORM APPEARS IN TEXT)
                        --------REGULAR (DIFF. FORM APPEARS IN TEXT)
                        --------TOTAL
                        --------TYPICAL
```

Figure 3. (Continued)

```
                        --------UNIVERSAL
                        --------WHOLE
                --------PLAN
                --------TACTICS
1508    THEORETICAL                                                     21
       *THEORETICIAN
       *        --------THEORY                                         36
                        --------ASSUMPTION
                        --------IDEA
        THEORETICIANS
1509    THEORIES                                                        36
        THEORIST
       *THEORY
                --------ASSUMPTION
                --------IDEA
1593    UNIFIED                                                         78
        UNION
                --------ACCORD
                --------AGREEMENT
                --------DISAGREE (DIFF. FORM APPEARS IN TEXT)
                --------SOLIDARITY
                --------UNDERSTANDING
        UNITED
        UNITE
        UNITS
        UNITY
1633    WARFARE                                                         293
        WAR
                --------AVIATION
                --------BALLISTICS
                --------BATTLEFIELD
                --------BATTLE
                --------BELLIGERENT
                --------CAMPAIGN
                --------COMBAT
                --------CONFLICT                                       24
                        --------BATTLE
                        --------DISAGREEMENT
                        --------OPPOSE (DIFF. FORM APPEARS IN TEXT)
                        --------WAR
                --------FIGHTING (DIFF. FORM APPEARS IN TEXT)
                --------GUNPOWDER
                --------HOSTILITIES
                --------MILITARY                                       285
                        --------FIGHTER (DIFF. FORM APPEARS IN TEXT)
                        --------GENERAL                               53
                                --------COMMON
                                --------NATURAL
                                --------NORMAL (DIFF. FORM APPEARS IN TEXT)
                                --------REGULAR (DIFF. FORM APPEARS IN TEXT)
                                --------TOTAL
```

Figure 3. (Continued)

```
                                --------TYPICAL
                                --------UNIVERSAL
                                --------WHOLE
                        --------MAJOR
                        --------MARSHAL
                        --------OFFICER
                        --------SOLDIER
                        --------WARFARE
                        --------WAR
                --------PACIFIC
                --------PEACE
                --------STRATEGY                                          283
                        --------GENERALSHIP (DIFF. FORM APPEARS IN TEXT)   53
                                --------COMMON
                                --------NATURAL
                                --------NORMAL (DIFF. FORM APPEARS IN TEXT)
                                --------REGULAR (DIFF. FORM APPEARS IN TEXT)
                                --------TOTAL
                                --------TYPICAL
                                --------UNIVERSAL
                                --------WHOLE
                        --------PLAN
                        --------TACTICS
                --------TACTICS
WARS
WARTIME
```

Figure 3. (Continued)

A quick inspection of this output shows that the format differs slightly from that for Hamlet. Before proceeding to a discussion of the output's contents, a brief description of the changes in format seems desirable. First, in this instance only, the use of the asterisk is different. Here, with the exception of the words ALL, THEORETICIAN and THEORY, words preceded by asterisks are words which do not occur at all in the RAND translation; these words have never qualified as primary words but are of interest because of their potential for distinguishing between the two translations (see pp. 68-71 of this report). Words in this category have been included only for the alphabetical range a - d. ALL appears in both translations but

does not have primary frequency in the Praeger edition. THEORETICIAN and THEORY occur in both translations but never qualify separately as primary; together, however, they do have primary frequency and, because of their related meanings, lists were compiled for both words. Second, the notation DIFFERENT FORM APPEARS IN THE TEXT is introduced in this output. This notation has been introduced to take care of the problem of cognates. As an example, in the list of words associated with AIM, the word ASPIRE does not occur but a cognate, ASPIRATIONS, does; the word INTEND does not occur but INTENDED does. In most cases, the difference in form has little or no effect on content analysis (the majority of the cases are occasioned by a variation in tense), but in some few instances the difference in form does seem important (for example, although PRINCIPAL and PRINCIPALITY do seem appropriately grouped together in some cases, in others they do not--hence, it is well to know which form actually occurs). The researcher can quickly determine which cognates do occur by glancing at the complete output (not shown in Figure 2) for this chapter. We could also print out the cognates and we may well add that feature to the program when making the programming transition from the Philco 2000 to the IBM 360 system.[18] Different forms resulting from "s" "'s" and "s'" are ignored. A glance at the output including ARMAMENT shows a slightly different procedure when the

[18]The Philco 2000, which has been the computer used by this project, is being replaced by models in the IBM 360 system. It seems advisable to wait on minor programming changes until the recoding process is under way.

difference exists between the thesaurus primary word entry (that saved by the computer) and the words within the primary root group. The root group including ARMAMENT, ARMAMENTS, ARMED, ARMIES, ARMS, ARMY'S, and ARMY does not have the exact thesaurus form, ARM. Since the program has already printed out all the words in the root group, it seems sensible to introduce the notation DIFFERENT FORM APPEARS IN THE THESAURUS to indicate the existence of a different thesaurus form. Again, this form may well be printed out in a future version of the program.

To return to a more substantive discussion of Figure 3, the first characteristic of this output to notice is that some content word root groups occur with extremely high frequency. The root group containing WAR has 293 occurrences, that containing MILITARY, 285 occurrences, and that containing STRATEGY, 283 occurrences. Groups of more than 100 occurrences are those including the words POLICY and POLITICAL (159 occurrences), ARMS and ARMIES (144 occurrences), COUNTRY (122 occurrences), ECONOMIC (130 occurrences) and FORCES (105 occurrences). Even considering the fact that the total number of content word occurrences in Act V of Hamlet is just slightly more than one-third that in chapter one of the Praeger edition of Sokolovsky (3308 and 9734, respectively), the difference between these texts is indicated by the fact that in Act V of Hamlet the two root groups with the highest frequency (containing GO and NO) have only 32 and 30 occurrences, respectively. There are only 5 root groups with occurrences of 20 or more; these are the groups containing HERE'S (27 occurrences), MORE (24 occurrences), LET (23 occurrences), DO (22 occurrences) and GOOD (20 occurrences). Furthermore, in all cases

excepting NO, GOOD and MORE (which is only marginally a content word), the words with the highest frequencies in Act V of Hamlet are verbs which may well be connected with the mechanics of action required for staging a play. In other words, in Hamlet the concept terms or major theme-bearing terms occur with relatively low frequency, indicating a much more varied, much less obvious style (including subject matter) than that in chapter one of Praeger. Thus, although the use of a high cut-off point (50 or more) for primary word designation in this translation enables us to study 18 root groups and their associated lists, this information is likely to seem obvious because of its very repetitiveness. This impression is an accurate perception of the structure of chapter one: there are a few major organizing concepts which dominate the chapter.

Among these concepts, the overwhelming emphasis upon root groups containing WAR, MILITARY, and STRATEGY has already been noted. The frequency of each of these groups is well over 100 occurrences more than that of the next most frequent primary root group--that containing POLICY and POLITICAL. The researcher might well assume that if strategy is important, then war and the military must be vital to that strategy. Or, if the strategy is military strategy, that strategy must focus heavily upon war. The sublists under WAR include MILITARY and STRATEGY and the sublists under MILITARY include WAR and STRATEGY. AVIATION is relevant to war, as is the science of projectiles (BALLISTICS). The latter might bring missiles to mind, and a glance at the sublist under ARMS would reveal the presence of MISSILE, ROCKET and NUCLEAR.

The information analyst should also notice that FIREARM and GUN appear on that list and that GUNPOWDER is on the list linked to WAR (which in turn is linked to ARMS, etc.). The terms cited in the above two sentences might imply two kinds of strategy, that involving nuclear missiles and that involving conventional weapons. Pursuing this lead further, the investigator could check the complete output and discover that words associated with NUCLEAR occur 30 times and the terms MISSILES and ROCKETS a total of 7 times. On the other hand, words associated with FIREARMS and GUNS occur a total of 14 times, and the word ARMED occurs in the phrase ARMED FORCES 47 times (the latter information can be obtained from a perusal of the indexes for ARMED and FORCES, or from a MAPTEXT printout, or by going directly from the index entries for ARMED to the indicated context). A further investigation of the contexts for the occurrences of ARMED FORCES would show that the phrase seems most often to refer to conventional warfare. Thus, so far as frequency is concerned, conventional warfare receives a greater emphasis than nuclear warfare (approximately 61 to 47), although the latter is by no means ignored. A complete reading of the chapter shows this picture to be accurate; there is, in fact, an ambivalence toward nuclear warfare, revealed in such passages as the following:

> The country that finds itself in a catastrophic situation as the result of mass nuclear-missile strikes may be forced to surrender even before its armed forces have suffered any decisive defeat. But we must remember that such results can be accomplished only by means of force, by means of armed conflict.[19]

---

[19] Praeger, p. 107.

Or, again, having earlier noted that nuclear warfare had completely changed military strategy, this statement is made: "Modern war involves mass armies..."[20]

War's antonym, PEACE, does occur, and a glance at the complete output would show that words having PEACE as a root have a frequency of 15. PACIFIC is a less definite indicator, since it can refer to a geographic area as well as to peaceableness. The complete output shows that PACIFIC occurs just once, and a glance at its context indicates that its meaning here is geographic. Therefore, it does not add extra emphasis to the slight weight on the side of PEACE; the frequency of WAR seriously overbalances any verbal considerations of PEACE in this chapter. Again, this frequency accurately reflects the content of the chapter, even though it concludes with a rather extended testimonial to the "peace-loving"[21] policies of the U.S.S.R.; the emphasis of the chapter is on military strategy rather than on strategies for peace.

After the major emphasis upon war, strategy, military, and arms, all of which seem to fit together thematically, the next important consideration, in terms of frequency, entails the economy, with which MONEY and TRADE are linked. The range of words in the ECONOMY root group (ECONOMIC, ECONOMICS, ECONOMIES, ECONOMISTS, plus ECONOMY) as well as the word TRADE linked to ECONOMY suggests a preoccupation of somewhat greater breadth than, for

---

[20] Praeger, p. 36.

[21] Praeger, pp. 47 and 48.

example, a relatively simple interest in how much is to be budgeted for arms. Supporting this contention are such statements in the text as the following:

> Economic conditions determine not only the issue of war in general, but also the operational method of the army--i.e., the method and means by which war is conducted, its strategy.[22]

By looking at the complete VIA output for this chapter, the information analyst would see that the root groups including CAPITALISM and SOCIALISM have a rather high frequency (35 and 60, respectively), and he might suspect that the implications for strategy of these two economic systems are considered.

Among other high-frequency root groups, those including DEVELOP and POLICY, with a frequency of 73 and 159, respectively, seem germane to the emphasis upon strategy and also upon AIM, another word with a large number of occurrences. If the cut-off point for primary words were dropped to 40 or more occurrences, two more root groups--those including METHODOLOGY (45 occurrences) and SCIENCE (42 occurrences)--related to the general cluster of STRATEGY and POLICY would be apparent; the focus in this chapter, so far as strategy is concerned, is with a science or methodology which will permit military success. Often, science is so broadly defined as to include, for example, a socialistic economic system as part of the scientific methodology. On other occasions, it refers to technological development, as it affects strategy.

---

[22] Praeger, p. 27.

Another root group which has a high frequency is that containing MORALE. Again, a look at the complete chapter reveals that MORALE is given detailed attention as a factor in war; its designation as a primary word is, therefore, an accurate picture of its conceptual importance.

Of the other primary root groups, those including SOVIET, SOCIALIST, STATE and COUNTRY all seem related. It might be profitable to perform a more detailed examination of the terms STATE and COUNTRY than the VIA run shown in the output provides in order to see whether STATE is more closely tied to a system of government and COUNTRY to a given geographical area--in other words, to see whether STATE is more politically oriented as a term and COUNTRY more nationalistically oriented. If this were the case, the fact that country has a higher frequency might well be of interest. The current output is not dense enough to answer this question, although the presence of PATRIOTIC on the list headed by COUNTRY may be suggestive.[23]

The one root group among these primary words which may seem to contain a curious combination of words is that containing both POLICY and POLITICAL. An argument can be made both for leaving them together and for separating them into distinct root groups; in the future, we may well separate them but in this particular instance their common grouping does accurately reflect the content of the chapter. POLITICAL and its cognates occur 86 times, while

---

[23] This theory re the possible distinction between STATE and COUNTRY might at first seem untenable because each term appears on a list headed by the other. However, the fact that they are in many senses synonymous cannot be denied; the interesting question is whether the lists associated with them also reveal the senses in which they are not synonymous.

POLICY and POLICY'S occur 73 times. In many cases, the terms are used almost interchangeably. In the case of the root group just mentioned, as in a number of others, the primary words with lesser frequency are related to those, such as STRATEGY, with dominating frequency.

The one other primary root group of special interest is that including GENERAL. As the words linked to GENERAL suggest, its primary use in this chapter is to refer to the universal, typical, or common. It is entirely appropriate that the one word among the primary root groups we've considered that seems in the main a modifier of other, more specific, terms, is GENERAL. In translation, the name of the chapter is General Concepts; the chapter is given over to a very general consideration of the impact of the economy (including economic systems), of political structure and diplomatic policy, of morale, and of technology upon military strategy and the wars which implement that strategy. VIA reveals these emphases, as well as others--such as that upon conventional warfare--which are not immediately obvious from the structure of the chapter or from a quick reading of it.

For a more subtle, detailed analysis, the primary word cut-off point could be lowered, and groups of sublist words could be used as additional points of departure. In addition, the output from this chapter could be compared with the output from subsequent chapters and new runs could be made on this chapter in the light of information gleaned from subsequent chapters. (For Hamlet, a separate thesaurus was compiled for each act and computer output obtained, after which runs using the complete thesaurus were made on each act; this procedure very clearly showed both continuities and

discontinuities.) To begin with, however, the rather thin output reproduced in Figure 3 would doubtless indicate to the researcher whether or not he wanted to examine the chapter in greater detail.

So far as content analysis is concerned, VIA's output seems useful and flexible. Hamlet and Sokolovsky are certainly very different types of text, and the researcher's interest is not necessarily the same in both cases. For example, no one is likely to use VIA on Hamlet to determine whether or not he wants to read Hamlet. Rather, VIA might be used to provide a more detailed basis for interpretation or to make a more general study of the genre of tragedy. So far as documents like Sokolovsky's are concerned, VIA could be used both for gross information retrieval purposes (i.e., should the document be designated for close examination?) and for close analysis. The man-machine interaction which VIA provides can be used both to guide close reading to important sections of a document and to enrich comprehension of that reading so as, for example, to speed effective response to emergency diplomatic communications from other states. For this range of purposes, VIA seems to show considerable promise.

2. Semantic Analysis. VIA provides various sorts of aids for the study of semantics. First, the grouping of words by root and by lists and the linking of lists to other lists pulls together into rather easily grasped displays a good deal of material for the perusal of the semanticist. The retention of "redundant" lists linked back by way of varying paths to the

primary root group adds to the richness of the semantic information.[24] Second, the detailed study of words in context required before some lists of possible-associated-words are compiled and before some generalizations about VIA's output can be made[25] contributes to the general understanding of patterns of word usage which semanticists must have in order to construct usable models for all types of automatic language processing.

A good example of the retention of redundant lists occurs in the output for the first chapter of the RAND translation of Sokolovsky--for the root group containing the word DEVELOP. This output is reproduced in Figure 4.

```
400   DEVELOPED
      DEVELOPING
      DEVELOPMENT
      DEVELOP
                        ADVANCE
              --------AGE
              --------AUGMENT
                                ACCUMULATED
                                ADDED
                                EXPANDED
                                INCREASED
                                LESSENED
                                REDUCED
                      --------REINFORCED
      --------DECLINE
                                ADD
                      --------AUGMENT
                                        ACCUMULATED
                                        ADDED
                                        EXPANDED
                                        INCREASED
                                        LESSENED
                                        REDUCED
                              --------REINFORCED
```

Figure 4. VIA Output for DEVELOP Root Group, RAND Translation

---

[24] The term "redundant lists" refers to lists comprised of identical words and therefore redundant so far as content analysis is concerned.

[25] See discussion on pp. 61-66 of this report.

```
                                DEVELOP
                                GROW
                                INCREASE
                                LESSEN
                                DECREASE
                                        ADD
                                --------AUGMENT
                                                ACCUMULATED
                                                ADDED
                                                EXPANDED
                                                INCREASED
                                                LESSENED
                                                REDUCED
                                        --------REINFORCED
                                --------DECLINE
                                        DEVELOP
                                        GROW
                                        INCREASE
                                        LESSEN
                        EVOLUTION
                        EVOLVE
                        PROGRESS
                --------RIPEN
        DEVELOPS
```

Figure 4. (Continued)

Most prominent among the lists repeated in this output is the list for which ACCUMULATED is the first word. This list is always linked back to DEVELOP through AUGMENT, but the linking pattern involving AUGMENT varies. The first time AUGMENT occurs, it is directly linked to DEVELOP. The second time, it is linked to DECLINE, which is linked to DEVELOP. And the third time, it is linked to DECREASE, which is linked to DECLINE, which is linked to DEVELOP. For semanticists who are interested in developing lattices and other types of networks which will enable them to assign meanings to given words, VIA's display of intersections and subconnections should be very useful. As for the lists of words in which AUGMENT occurs in this output, two of them are similar, although not identical, and one is quite dissimilar.

In the list directly linked to DEVELOP, AUGMENT is one of a number of synonyms--ADVANCE, AGE, EVOLUTION, EVOLVE, PROGRESS, RIPEN--which support and add ramifications to the general concept of DEVELOP and DEVELOPMENT. In the list linked to DECLINE, AUGMENT is an antonym to DECLINE which, in turn, is an antonym to DEVELOP. AUGMENT is again an antonym in the list linked to DECREASE, but DECREASE is a synonym for DECLINE, which in turn is an antonym for DEVELOP. The two lists on which AUGMENT is an antonym are very similar, the only difference being provided by the presence of DECREASE on the list headed by DECLINE in one case, and by DECLINE on the list headed by DECREASE in the other. The input lists of possible-associated-words for DECLINE and DECREASE obviously contain a number of words in common but they are not identical. Several words associated with the concepts of decadence and deterioration are associated with DECLINE but none of these words occurs in the text being examined. Thus, the similarity of these lists is of interest not only to the semanticist but to the information analyst as well, who can conclude that the concepts of DECREASE and DECLINE are used almost interchangeably in this first chapter of the RAND translation.[26]

Examination of the occurrences of primary words in context for both translations of Sokolovsky leads to some observations which are again of interest to both the semanticist and the information analyst. Among those root groups qualifying as primary in frequency, those groups containing the

---

[26] Early Marxists argued that all qualitative statements are really quantitative statements. The interchangeable usage of DECREASE and DECLINE in this chapter may indicate the continuing strength of that argument in the U. S. S. R. today.

words AIM, ARM, DEVELOP, FORCE, GENERAL, OPERATION, and STATE provide the greatest possibility for ambiguity in function and meaning and thus for a misleading interpretation of the output. In order to explore the actual situation, a brief look at each of the words seems in order.

Although AIM can be used both to describe goals or purposes and to describe the action, either specific or general, of aiming, this disparity of function would not seem to result in a very serious ambiguity so far as content analysis is concerned. The act of aiming implies a goal or purpose just as a goal or purpose implies some sort of direction in the actions connected with the goal or purpose. Nonetheless, a glance at the words in the root group and their contexts in the chapter shows that there is no need to worry about this ambiguity of function whatever its implications for meaning might be. In both translations, the words AIM, AIMS, and AIMED occur. The group has a frequency of 54 in the RAND translation and 52 in Praeger. AIMS occurs 39 times in both translations, AIM occurs 12 times in RAND, 11 times in Praeger. Neither is used even once as a verb. Thus, in 50 of 52 cases in Praeger, and 51 of 54 in RAND, the words in this root group serve the same function. This rigidity of function, if characteristic, is an important guide to style, both for purposes of content analysis and for identification of style on the generic, if not the individual, level. In the other two cases in Praeger and three cases in RAND, the form AIMED signals the past tense of a verb.

The root group containing ARMS also includes such words as ARMAMENTS and ARMIES. In the abstract, this grouping might appear potentially

misleading, but in fact it is not. The word ARMS, occurring only three times in each translation, is always used synonymously with armaments; it never, for example, designates a branch of authority or, much less likely, a part of the human anatomy. It might seem foolish to raise the latter possibility when dealing with a book on military strategy, but semanticists and linguistic purists frequently insist on raising all possibilities when indicating the pitfalls awaiting the information analyst. The important point of this example and others which will be cited is that there is semantic rigidity in these instances. The function of the word ARMED in this group is also interesting. Although it might serve as either a verb or an adjective, in fact it never serves as a verb. Its 71 occurrences in Praeger and 73 in RAND are always adjectival, in phrases such as ARMED CONFLICT or ARMED COMBAT. Again, as in the case of AIM, there is rigidity of function as well as exclusion of the verbal function.

Even though DEVELOP is unequivocally a verb, the emphasis in the root group containing DEVELOP is upon the noun, DEVELOPMENT. Fifty-two of the 73 occurrences in this group in Praeger are of DEVELOPMENT; 49 of 73 in RAND are DEVELOPMENT. The verbs DEVELOP and DEVELOPS occur 15 times in Praeger and 13 times in RAND. DEVELOPED is used both as a verb and adjectivally. As was true in the case of AIM, these differences in syntactic function do not have major implications so far as content analysis is concerned because one form implies the other.

FORCE, FORCES, and FORCED, all of which occur in the same root group, are again words which can fill several syntactic functions. However, the same lopsidedness in favor of nominal use apparent in earlier groups is evident in this root group. FORCE occurs 21 times in Praeger and 16 times in RAND; in each translation, it is used as a verb just once and that is in a quotation from another work. FORCES is used 76 times in Praeger and 78 times in RAND; again it functions as a verb just once in each case and in Praeger it again serves this function within a quotation. FORCED, appearing twice in RAND and seven times in Praeger, is used verbally (predominantly passively--six times) in every case. So far as the subtleties of meaning are concerned, FORCE can be quite abstract or it can be a bit more concrete, as when it implies a military unit of some sort. Again, as in earlier cases, one meaning implies the other. In chapter one of Sokolovsky there is the same single-mindedness so far as meaning is concerned as there is so far as syntactic function is concerned. The word FORCES is almost always used to designate a body of armed men; in fact, on 66 occasions in Praeger and 78 in RAND, FORCES occurs in a phrase such as ARMED FORCES. FORCE occurs in a similar context nine times in Praeger and seven times in RAND. Although the words linked to the FORCE root group in VIA's output are conceptually meaningful and accurate, the information analyst might wish to add words to the thesaurus input which would provide a visual connection between force and army and war (e.g., adding FORCE to the thesaurus lists under ARMY and WAR). For our experimental runs, we attempted to avoid such special-purpose inputs; however, part of our effort to automate

the manual search will entail consideration of co-occurrence, and it is likely that a future version of VIA will automatically take care of the link between FORCE and ARMED, etc.

The word GENERAL is another interesting case of a word which has several meanings. One might anticipate that general would designate a military officer, for even though general is not a title used for Russian military officers, U. S. military figures are mentioned with some frequency. This, however, is not the case. In Praeger, GENERAL occurs 42 times but only 4 of those occurrences specify a military officer; 3 of them are in titles ("General Motors" occurs twice) and in the other 35 instances GENERAL means common, universal, etc. The same emphasis is found in RAND: of a total 45 occurrences, 40 are used in the universal sense.

The root group containing OPERATION and OPERATIONS includes no verb form, not even the slight representation by a passive, past tense as in some of the other groups being discussed. Further, whether OPERATION or OPERATIONS is under discussion the focus is always military; the purist might wish to raise the possibility of OPERATION being used, for example, to describe the function of equipment or even to describe surgical procedures. Those possibilities are not, in fact, relevant. Again, there is semantic rigidity. There are 54 occurrences of OPERATION and OPERATIONS in Praeger and 59 in RAND. Often, the phrase MILITARY OPERATION or MILITARY OPERATIONS is used.

STATE and STATES, in the last root group being considered in this discussion, are both syntactically and semantically ambiguous. So far as syntax is concerned, in RAND, STATE and STATES never function as verbs; in 118 instances, they serve either as nouns or adjectives; in two instances STATE appears in the title, "Secretary of State." In 78 instances in Praeger, STATE and STATES function as nouns or adjectives; in one instance STATES is used as a verb. As to meaning, STATE can of course refer to something roughly approximating condition, or to a country. The latter meaning is the top-heavy favorite in the translations of Sokolovsky. In RAND, given 66 possibilities, STATE means condition just 5 times; in Praeger, given 47 possibilities, STATE means condition just 4 times. The other 61 and 43 occurrences refer to the political unit called a state. STATES is not used once to imply conditions. In RAND, 21 of its 52 occurrences refer to political states and 32 occurrences are part of the title, "United States." In Praeger, 9 of 35 occurrences refer to political states and 26 occur in the title, "United States."

These observations support speculations by semanticists such as Margaret Masterman and Karen Sparck Jones: in a text which might be labeled "platitudinous," a word tends to be used in the same way, so far as meaning is concerned, throughout the text. The importance of this tendency both for content analysis and information retrieval is considerable if a large number of documents turn out to be on the platitudinous side. If this were the case, there would not need to be a great deal of anxiety as to whether the content analyst were being misled by the apparent frequency of a given

word when, in fact, the word represented several different concepts. Semantic rigidity would ensure that it did not represent several different concepts. Or, in the case of automatic information retrieval, the establishment of the appropriate semantic frame for one occurrence might make possible a considerable reduction in the computer time devoted to analysis of later occurrences. Many documents in which an information analyst would be interested might well fall into this platitudinous category.

Consistency of syntactic function is also of importance, both for content analysis and information retrieval, and for identification of individual, generic, or cultural style. So far as content analysis and information retrieval are concerned, syntax often provides clues as to meaning; consistency of syntactic function will make for a surer, probably quicker, guide. So far as identification of style is concerned, consistency of syntactic function may indicate distinctions which are useful, for example, for clearer understanding of another culture and for communication with that culture. Using the translations of chapter one of Sokolovsky as a case in point, if the emphasis there upon nouns and adjectives (including verbs used nominally--gerunds--and adjectivally) rather than upon verbs is true to the original (and the fact that the same emphasis is present in both translations suggests that it may be), and if this chapter is in any way representative of a more general cultural emphasis, then this fact might be of importance to the syntactic emphasis we use in, for example, diplomatic documents. Too great an emphasis upon verbs, often implying action, might be viewed with alarm, even suspicion. It may be that we talk the language

of other nations even less effectively and less meaningfully than we suppose.

3. Stylistic Discrimination. So far as stylistic discrimination is concerned, VIA has been very little tested. In Figure 3, the partial output from chapter one of the Praeger edition, words preceded by asterisks are words which do not occur in the RAND translation. A similar output exists for the RAND translation (see Appendix, which gives the complete output for the alphabetic range, a - d). In an effort to see whether these differences represent real conceptual disparities between the two translations, as opposed to differences between translators' preferences as to word choice, all words linked to the asterisked words have been checked to see if they also are omitted from the other translation. The results of this investigation indicate that the disparity is not basically one of concept but rather is the choice of approximate synonyms within concepts.

Within the range defined above, (words in the alphabetic range a - d in chapter one) the Praeger edition includes 39 words not found in the RAND translation, and the RAND translation includes 32 words not found in Praeger. In the Praeger edition, 237 word tokens (as opposed to types, using the Herdan type-token distinction)[27] are linked to the 39 asterisked words. Only 16 of these tokens (representing 13 types) are of words which do not occur in the RAND translation. For the RAND translation, 233 word tokens are linked to the 32 asterisked words; 55 of these tokens (representing 26 types) do not

[27] Gustav Herdan, Type-Token Mathematics, Mouton and Company, 1960.

occur in the Praeger edition. In every case in Praeger in which words are linked to an asterisked word, at least one of the linked words occurs in RAND; this statement can also be made of RAND, with the exception of HOT which is linked as an antonym to COLD--neither word appears in Praeger, and since no other words are linked to COLD, this particular group is absent from Praeger. Thus, on the basis of this investigation, it seems safe to say that the significance of the differing word choices in the two translations is <u>not</u> one of concept but rather of translators' preferences for particular words.

So far as difference in style is concerned, the statistics above might suggest that the RAND translators are a trifle more inventive when it comes to using synonyms within a concept group (26 types which do not appear in Praeger as opposed to 13 types in Praeger which do not appear in RAND). Exploration of asterisked words for the alphabetical range e - z would provide a clearer picture of the word choice aspect of style in the entire chapter, and similar research on the other chapters in both texts might reveal either a more general pattern or recognizable variations within the pattern caused by the work of different translators on different chapters. We plan to explore these possibilities. As to the choice of individual words (as opposed to words within groups), only one of the asterisked words occurs with significant frequency. That word is CONFLICT which appears 24 times in Praeger and, of course, not at all in RAND; the RAND translators preferred COMBAT which occurs about 20 more times in RAND than in Praeger.

Of the other 38 asterisked words in Praeger, 32 occur just once, 3 occur twice, and 3 occur three times. As for the RAND translation, 25 of the 32 asterisked words occur just once, 5 occur twice, 1 occurs three times, and 1 occurs four times. In themselves, these occurrences are of no statistical importance; however, when other chapters are analyzed, the occurrences of these words might be grouped in various ways to discriminate between the translations on the basis of word choice. Ellegard used similar procedures when working with infrequently occurring words in an effort to identify the author of the Junius letters.

In addition to the interest of the asterisked words for the study of word choice, there are words and root groups which, although they appear in both translations, occur much more often in one than in the other. The group containing COMBAT, mentioned above, is an example of such a group. Research on these groups is planned, but it has not yet been undertaken.

The earlier discussion of primary words in Sokolovsky (see pp. 51-57) indicates that there is no major discrepancy between the two translations in this area. It would be extremely surprising if there were, since the primary word cut-off point was quite high in order to provide a picture of the major content or concepts of the chapter; if there were a major discrepancy here, the adequacy of either or both of the translations would be seriously questioned. Computer runs using a lower cut-off point may reveal some discrepancies, and experiments will be made with the cut-off point set at various levels.

Thus far, then, VIA as an aid to stylistic discrimination presents some interesting, even promising, possibilities, but much more research is needed before any conclusions can be drawn.

4. Plans for the Future. As it presently exists, VIA is a valuable, flexible tool for analysis and research. To increase its usefulness further, we plan to begin research toward automating the manual search now required for the initial input of possible-associated-words. Although the manual search is not impossibly time-consuming (two working days is a generous estimate for the time required to work out first-level lists for the primary words in Hamlet), it does somewhat slow down research using VIA. More important is the consideration that, although it is often desirable to be subjective when using VIA (if one is looking for a specific concept or reference there is no reason for not providing the appropriate verbal "leads"), there are many occasions when complete objectivity is required; even though it is possible to be quite objective by rigorously keeping to thesaurus and synonym dictionary entries, it is difficult to be completely objective; for example, sometimes a thesaurus entry seems so unlikely as not to be worth the trouble of putting it into the computer. Thus, it would be nice to retain manual search and input as an option, but to provide the capability of completely bypassing that search. Such capability is especially important if VIA is to be used for information retrieval or for the fast analysis of a great many documents.

Further automating VIA entails coping, in some measure, with the Scylla and Charybdis of every project aimed toward automating the production or analysis of language--syntactic and semantic descriptions of language. We have charted a number of courses for experimental runs; the results of these will determine the route we finally follow toward the goal of the automated search.

As mentioned earlier (p. 50 of this report), VIA will be reprogrammed this year for the SDC Research and Technology Division's new 360 complex. In the course of this reprogramming, a number of minor changes to the current program will be made. For example, if the thesaurus word and the cognate it retrieves from the text are not identical, both the message stating this fact and the cognate will be printed out; this addition to the program will obviate the need for the researcher to consult the output for the exact form occurring in the text. Attention will also be given to procedures for retrieving word pairs and triplets, when desired. Another useful improvement would be the inclusion of a concordance feature so that, for example, the contexts of primary words would be immediately available to the researcher. We also plan to provide an output comparison program, which will clearly show similarities and differences between word clusters VIA finds in the various sections of a given text or document. Now that VIA has shown itself a useful research tool, it seems worthwhile to refine and add to the output so that its usefulness will continue to be enhanced.

B. MAPTEXT

MAPTEXT is based on the premise that it is often desirable to look--with a certain controlled selectivity--at a text in some configuration other than the verbal form. Freed from semantic interference, patterns of word usage hitherto obscured may become apparent. Statistical studies have provided a popular mode of nonverbal representation and they form a part of the MAPTEXT "package." We believe, however, that a more easily visualized nonverbal representation is also desirable, and toward that end we have devised two procedures which "map" the verbal text into another, abstract format.[28]

MAPTEXT can be completely independent of VIA, or it can use any or all of VIA's output as input. In the short period during which MAPTEXT has been in use, it has been our experience that it quickly clarifies some aspects of VIA's output. For example, although VIA provides an index showing the location of each word occurrence, the perusal and collating of these indices takes time; even the concordances for specified words (such as primary words) that we propose to add to VIA's output will demand a mental collating effort of some intensity because of the semantic complexity the verbal context almost inevitably implies. But a glance at the MAPTEXT representation of the primary root groups in chapter one of the Praeger edition of Sokolovsky (see Figure 5) quickly shows in full detail that, as one would expect in a

---

[28] Jan LaRue, of New York University, has begun to use abstract representations of musical themes; he has suggested that these representations, on the analogy of MAPTEXT, be called MAPSCORE.

general way, certain of these primary root group words frequently appear together: MILITARY and STRATEGY, MILITARY and OPERATION, ARMED and FORCES are among the examples. In addition, this representation of the primary root groups delineated by VIA indicates grouping of occurrences of the root groups; the perception of such grouping could, again, be achieved through

KEY TO MAPTEXT REPRESENTATION

AIMED	1	FORCE	F	SOCIALIST	3
AIM	1	FORCES	F	SOCIALIZATION	3
AIMS	1	FORCIBLY	F	SOCIALIZED	3
ARMAMENT	A	GENERALIZATION	G	SOVIET	4
ARMAMENTS	A	GENERALIZATIONS	G	STATE'S	5
ARMED	A	GENERALIZED	G	STATED	5
ARMIES	A	GENERALIZE	G	STATEMENT	5
ARM	A	GENERALIZING	G	STATE	5
ARMS	A	GENERAL	G	STATES'	5
ARMY'S	A	GENERALS	G	STATESMEN	5
ARMY	A	MILITARILY	M	STATES	5
COUNTRIES	C	MILITARISTIC	M	STRATEGISTS	S
COUNTRY'S	C	MILITARIST	M	STRATEGICALLY	S
COUNTRY	C	MILITARISTS	M	STRATEGIC	S
DEVELOPED	D	MILITARY	M	STRATEGUS	S
DEVELOPING	D	MORALE	2	STRATEGY	S
DEVELOPMENT	D	OPERATIONAL	O	UNIFIED	U
DEVELOPMENTS	D	OPERATION	O	UNION	U
DEVELOP	D	OPERATIONS	O	UNITED	U
DEVELOPS	D	OPERATIVE	O	UNITE	U
ECONOMIC	E	POLICY'S	P	UNITS	U
ECONOMICS	E	POLICY	P	UNITY	U
ECONOMIES	E	POLITICALLY	P	WARFARE	W
ECONOMISTS	E	POLITICAL	P	WAR	W
ECONOMY	E	POLITICIANS	P	WARS	W
FORCED	F	POLITICS	P	WARTIME	W
		SOCIALISM	3		

Figure 5. MAPTEXT Representation for Primary Root Groups in Praeger Translation of Sokolovsky, Chapter 1.

The symbol "-" represents all nonprimary content words.

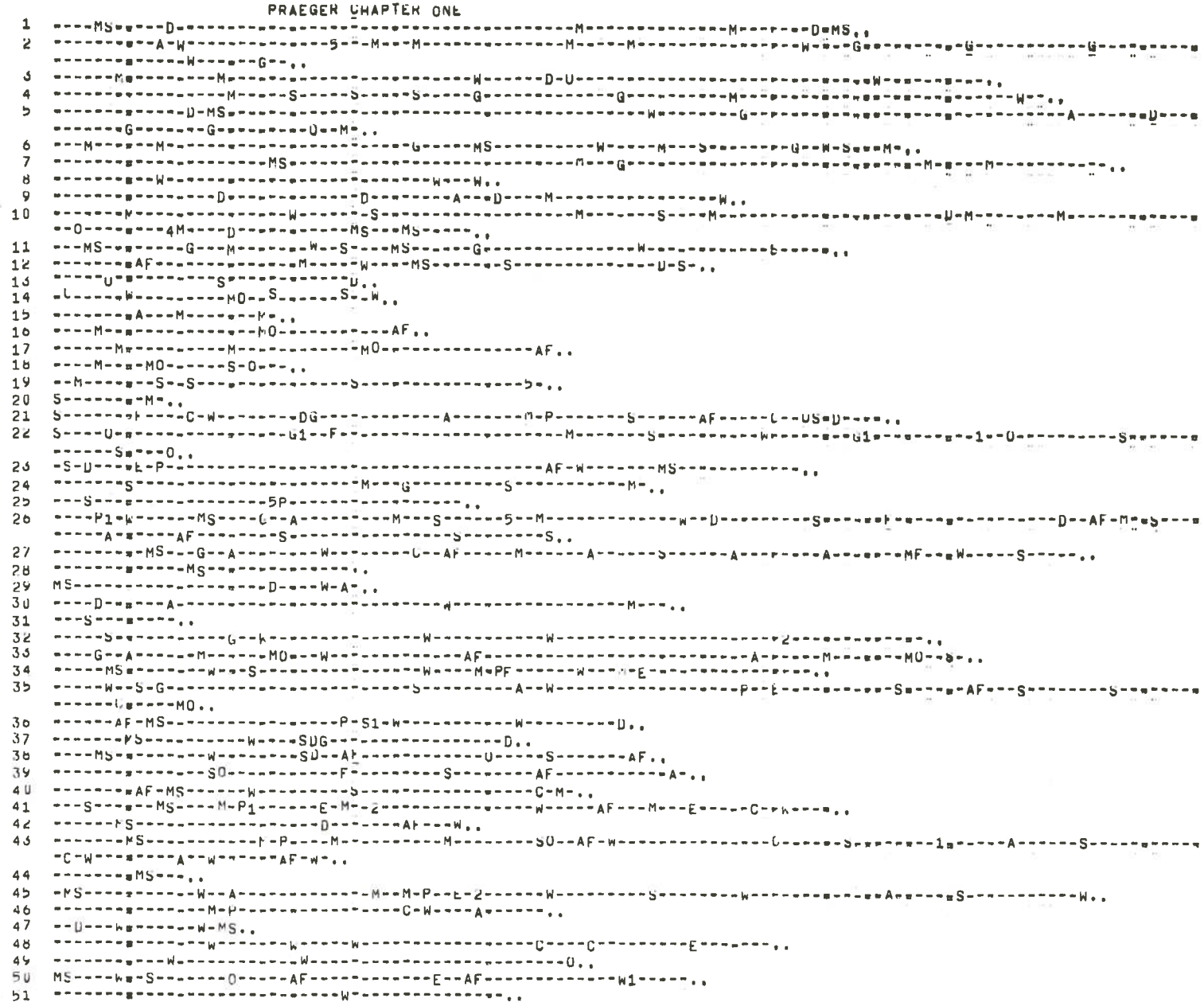


Figure 5. (Continued)

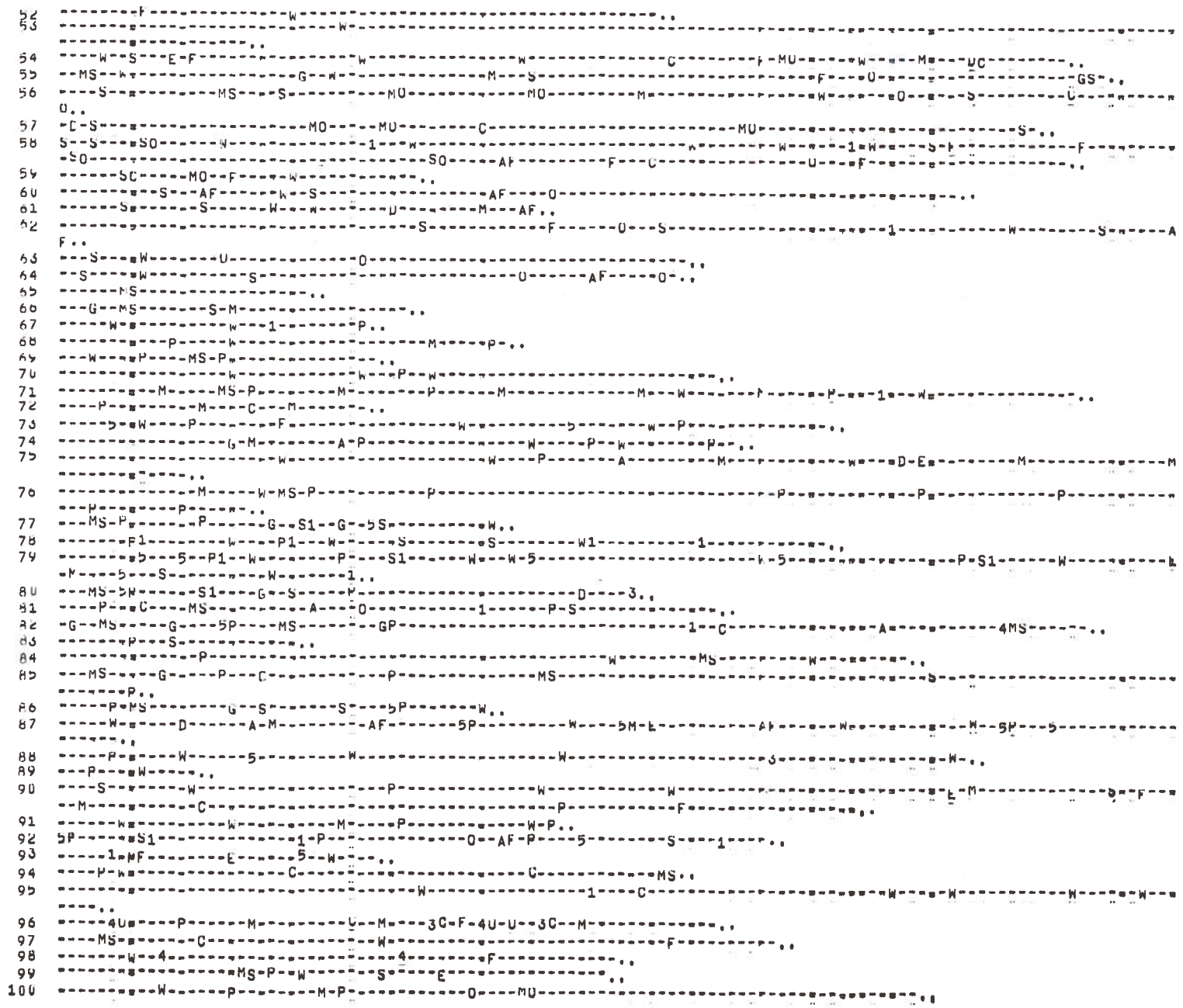


Figure 5. (Continued)

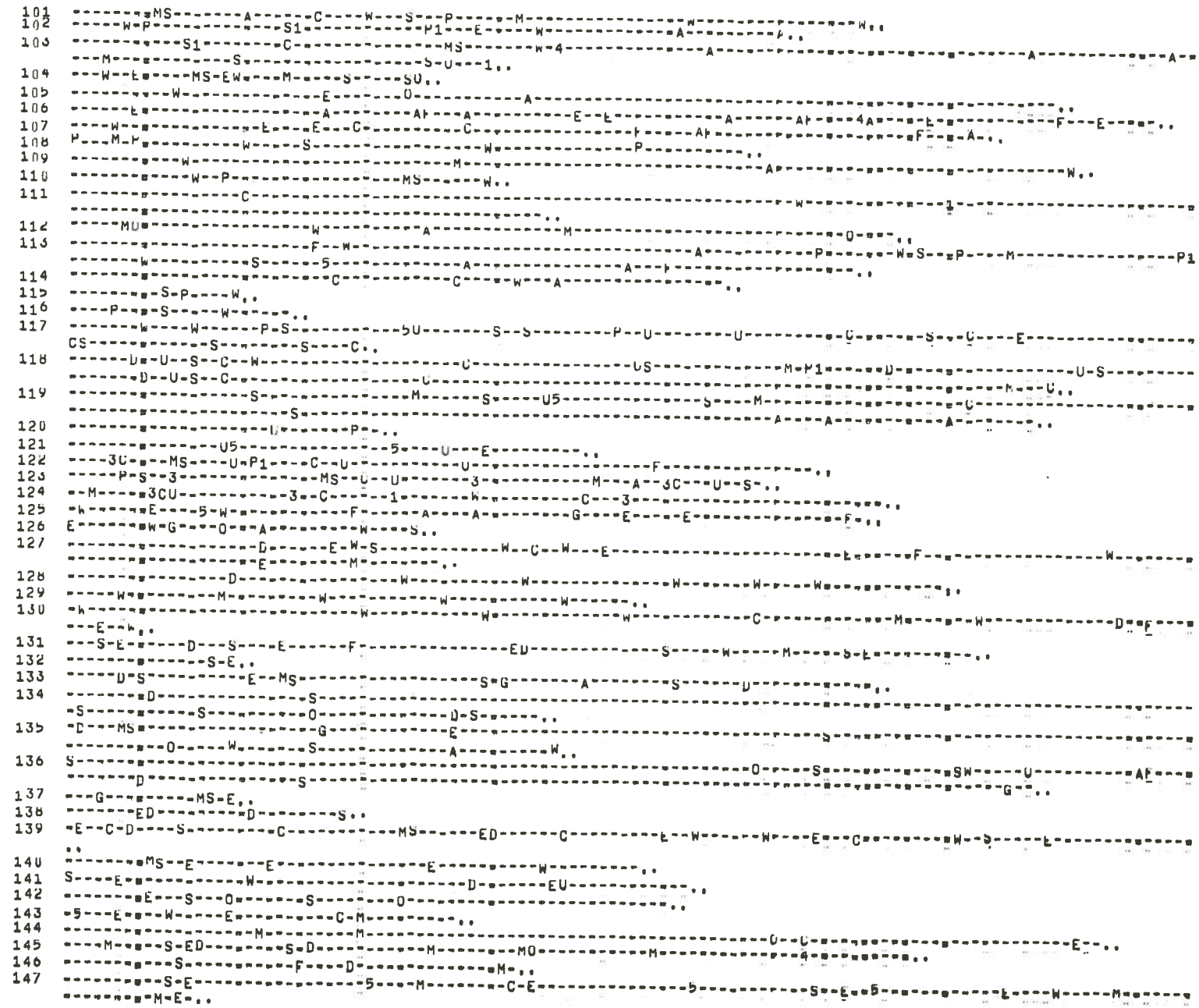


Figure 5. (Continued)

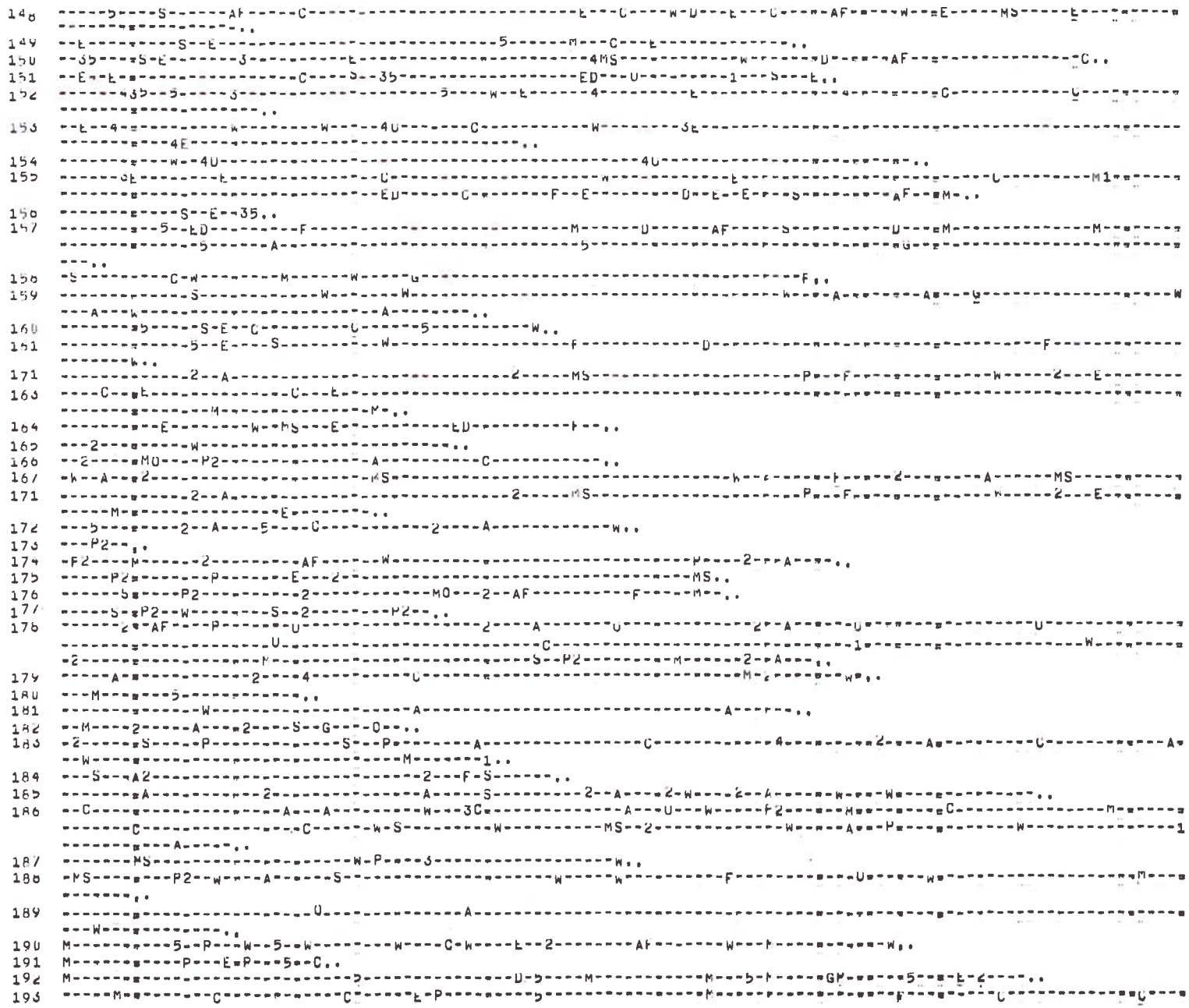


Figure 5. (Continued)

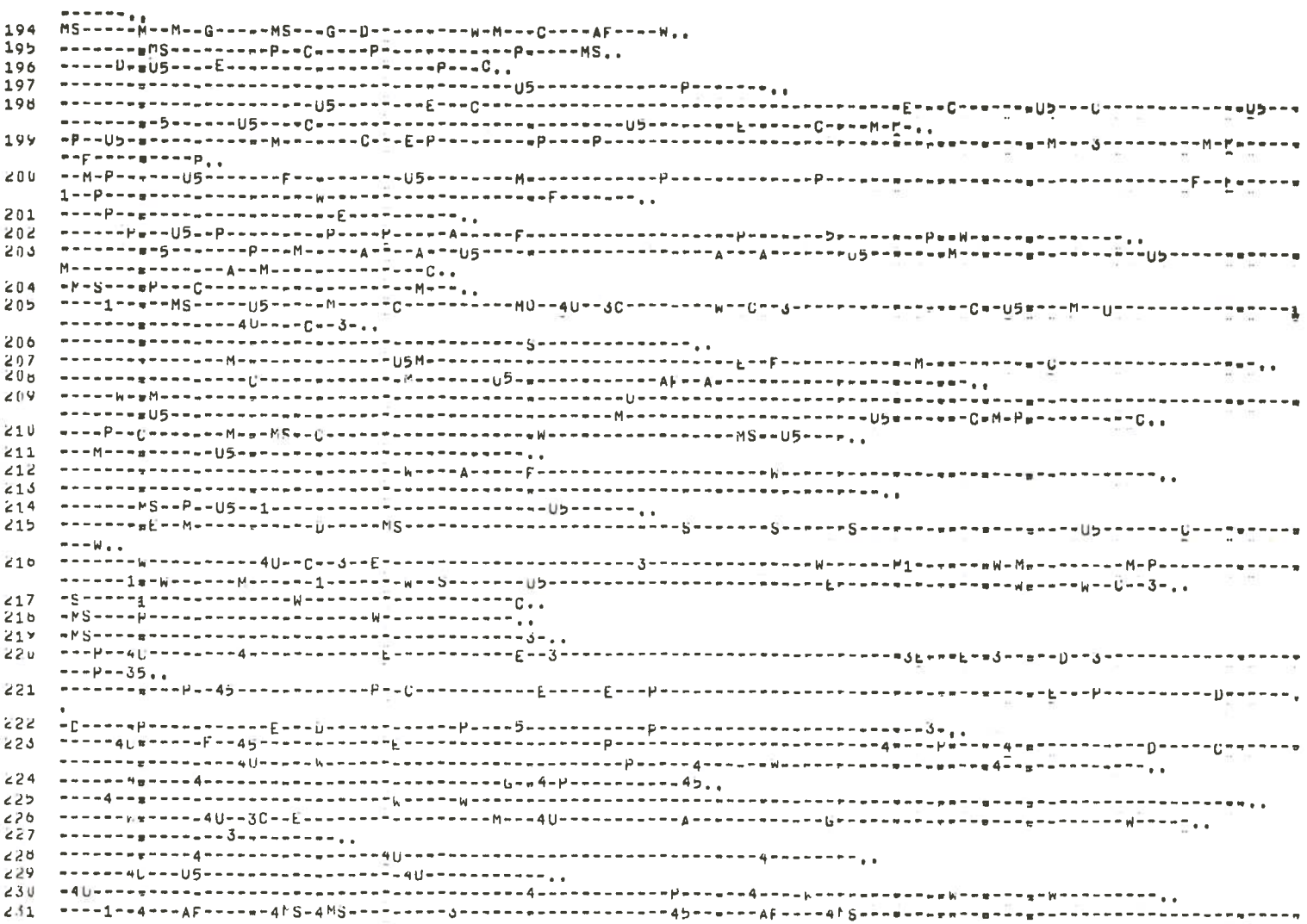


Figure 5. (Continued)

close perusal of the indices, but the MAPTEXT representation provides a much quicker guide to the possible clustering of concepts. Such a quick guide could be of considerable use, for example, when very fast response is required in diplomatic exchanges or other negotiations; the graph showing concept clustering could direct information analysts or diplomats to the sections of documents which would seem most likely to deserve close reading.

The use of MAPTEXT to explore the clustering of concepts has been strengthened through the addition of a second, condensed output shown in Figure 6.

4 *	48
8 ***	52
12	56 **
16	60
20	64
24 *****	68 *
28 **	72
32 *	76 *
36 **	80 ***
40 *	84 ***
44	88 **

Figure 6. Condensed Output Showing Occurrences of GENERAL in Praeger Translation, Chapter 1.

Numbers refer to paragraph intervals, asterisks to single occurrences. Thus, in paragraphs 1-4, GENERAL occurs once; in paragraphs 5-8, GENERAL occurs three times; etc.

92	164
96	168
100	172
104	176
108	180
112	184 *
116	188
120	192 *
124	196 **
128 **	200
132	204
136 ***	208
140 *	212
144	216
148	220
152	224 *
156	228 *
160 ***	231

Figure 6. (Continued)

The complete primary root group output in Figure 5 indicated that words in the root group of which GENERAL is a member (represented in that output by the letter G) seemed in some instances to be clustered. We wanted to see whether GENERAL, itself, seemed to be used in a clustered fashion because an

inspection of the words associated with it in VIA, as well as an inspection of the context of its occurrences, suggested the use of GENERAL to be so abstract as to be indicative, perhaps, of some broad stylistic habit rather than of a less generic stylistic pattern deriving from the consideration of a specific topic or idea. If the usage of this rather broadly defined term (sometimes used, for example, in the idiom "in general") seemed to be clustered, this clustering would support earlier observations about the rather artless rigidity characterizing the semantic and syntactic usage patterns of words such as STATE; that is, having fallen into a usage pattern, the writer tends to let it become something of a habit.

The condensed output in Figure 6, which shows occurrences of GENERAL per four-data-unit group, does indicate a tendency toward such clustering.[29] In order to check this clustering, we shall obtain MAPTEXT graphs on the other words VIA has linked to GENERAL. If the linked words do not cluster in the same location, but cluster elsewhere, the suggestion of artless rigidity may be supported; that is, once GENERAL is in mind, it is used for a while and then, UNIVERSAL having supplanted GENERAL, UNIVERSAL is used with more frequency, etc. If GENERAL and the words linked to it should cluster in the same locations, we would have to explore two possibilities: 1) the writing is not so rigid as supposed and the writer was indeed showing some variety in his word choice; 2) at least some of the words were more

---

[29] In this case, the data units are paragraphs. Data units may also be defined as sentences, stanzas, or whatever unit the researcher prefers to use.

specifically tied to a given topic or idea and we are dealing here with a specific concept rather than a passing taste for a certain set of rather vague modifiers and descriptors. None of these possibilities is mutually exclusive, but a more thorough exploration would probably establish the direction of emphasis.

For a less broadly defined term than GENERAL, MAPTEXT could be used to look for the clusterings of specific topics or ideas, or of groups of such topics/ideas. For example, using the VIA primary word output, it would be interesting to see if STRATEGY and PLAN are spatially reinforcing (clustered); or again, one might like to see whether ECONOMIC and POLICY, or MILITARY and POLICY, or all three of these terms cluster together. The graphic output would provide a quick guide as to the combinations the information analyst might especially wish to look into further. Separate MAPTEXT runs on any primary word and its synonyms would also quickly show whether a particular concept does indeed appear in restricted areas in the document or whether the synonyms occur elsewhere, providing for a more general use of the concept throughout the document or text than examination of the single primary word's graph would have led one to suppose.

The discussion of the representation of GENERAL indicates the value of the condensed graph output for the exploration of the subtleties of the stylistic patterns created by word choice. The correlation of such graphic output for a number of documents from the same field might also provide new leads for specialists working toward automatic abstracting, as to what areas of the documents are most likely to be richly loaded with information necessary to the abstracts.

Two statistical summaries are included in the current MAPTEXT output. The first of these, shown in Figure 7, indicates the number of occurrences per data unit of the specially sought linguistic units (either word(s) or punctuation marks). In the example, which is the summary for GENERAL, there were 197 data units in which the word did not occur at all, 27 data units in which it occurred once, 6 data units in which it occurred twice, and 1 data unit in which it occurred three times. Thus, this summary provides another guide as to the clustering of occurrences of the word.[30] The table shown in Figure 8 provides a summary of sequential occurrences of the specified linguistic units. The summary in Figure 8, for MILITARY and STRATEGY, indicates an isolated occurrence of one or the other of the words 301 times and a co-occurrence 80 times. Since the combinations MILITARY MILITARY, STRATEGY STRATEGY, and STRATEGY MILITARY, are unlikely unless there is intervening punctuation (which counts as a linguistic element and therefore would destroy any record of co-occurrence), it seems likely that the phrase MILITARY STRATEGY occurs 80 times.

0	197	0
1	27	27
2	6	12
3	1	3
	TOTAL	42

Figure 7. Totals of Occurrence in Data Units

[30] The column on the right in this table keeps track of the number of occurrences of the linguistic units being described by this table. For example, GENERAL occurs twice in 6 data units; therefore, the total number of occurrences in this category is 12.

1	301
2	80

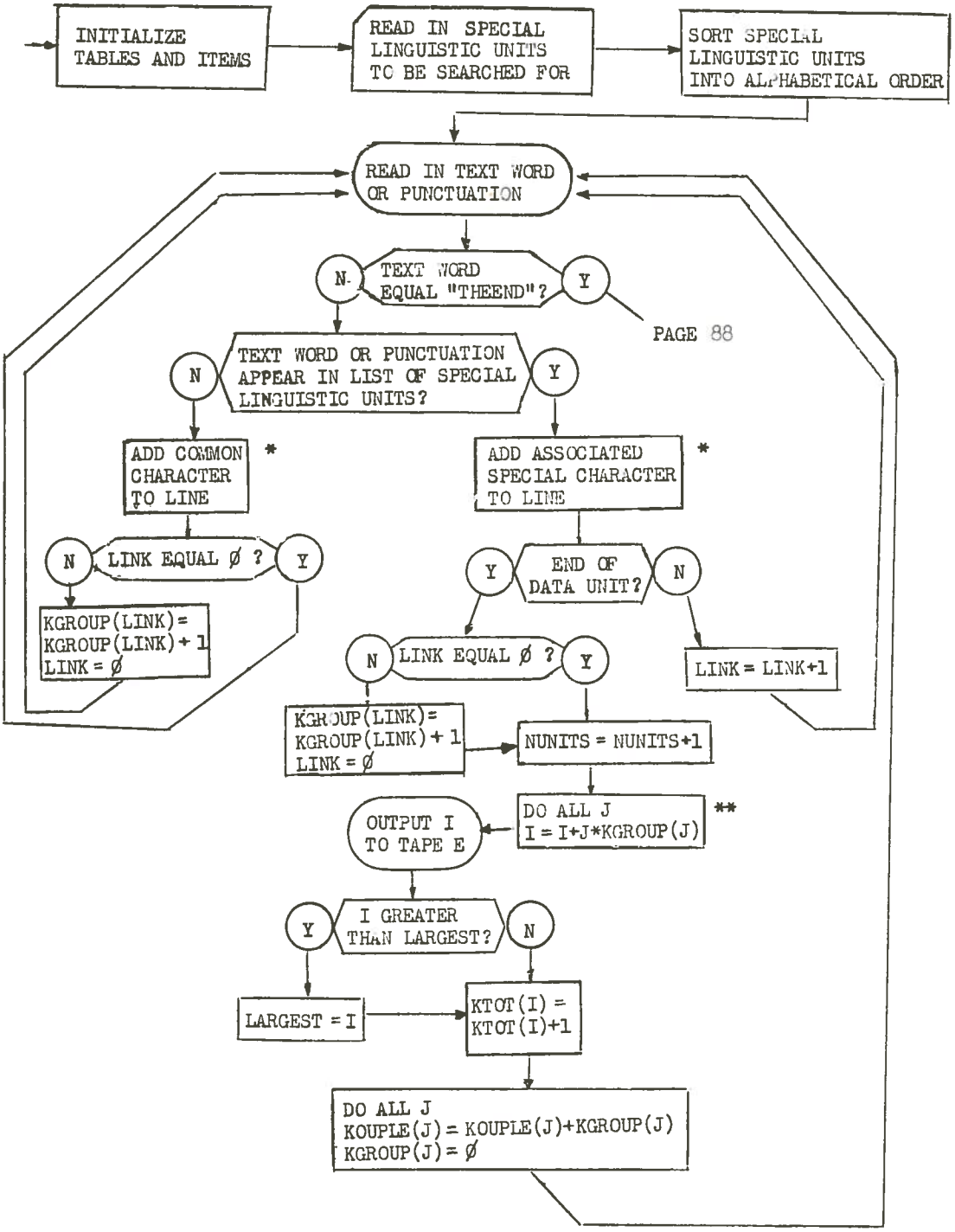
Figure 8. Number of Sequential Occurrences

As a program, MAPTEXT is open-ended. Its development has just begun, and we plan to add more graphic and statistical procedures as their need becomes apparent. Very likely, it will become important to build decision-making capacities into the program so that, at least generally, the computer, itself, will choose the graphical and statistical output procedures which would seem most informative in any given case. In time, the combined VIA/MAPTEXT programs should provide a very extensive, even comprehensive picture of the style of any given document or text; even in their current stage of development, they have greater analytical power than the many programs which have been aimed either at content analysis or statistical summation. The combination of these approaches, plus the addition of the graphic representation, represents a major advance toward seeing a style "whole" as well as in its disparate parts.

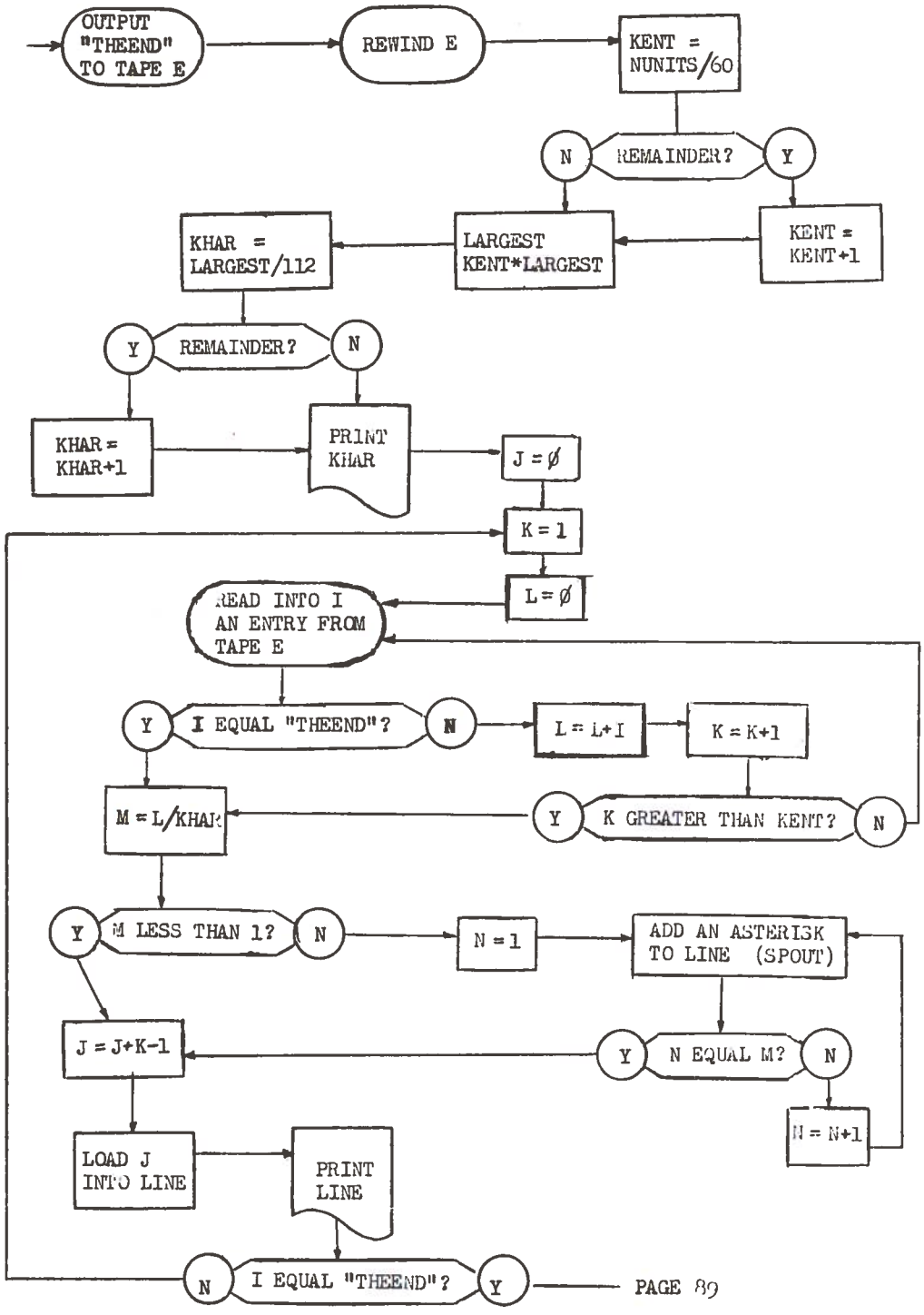
III. MAPTEXT: Program Description

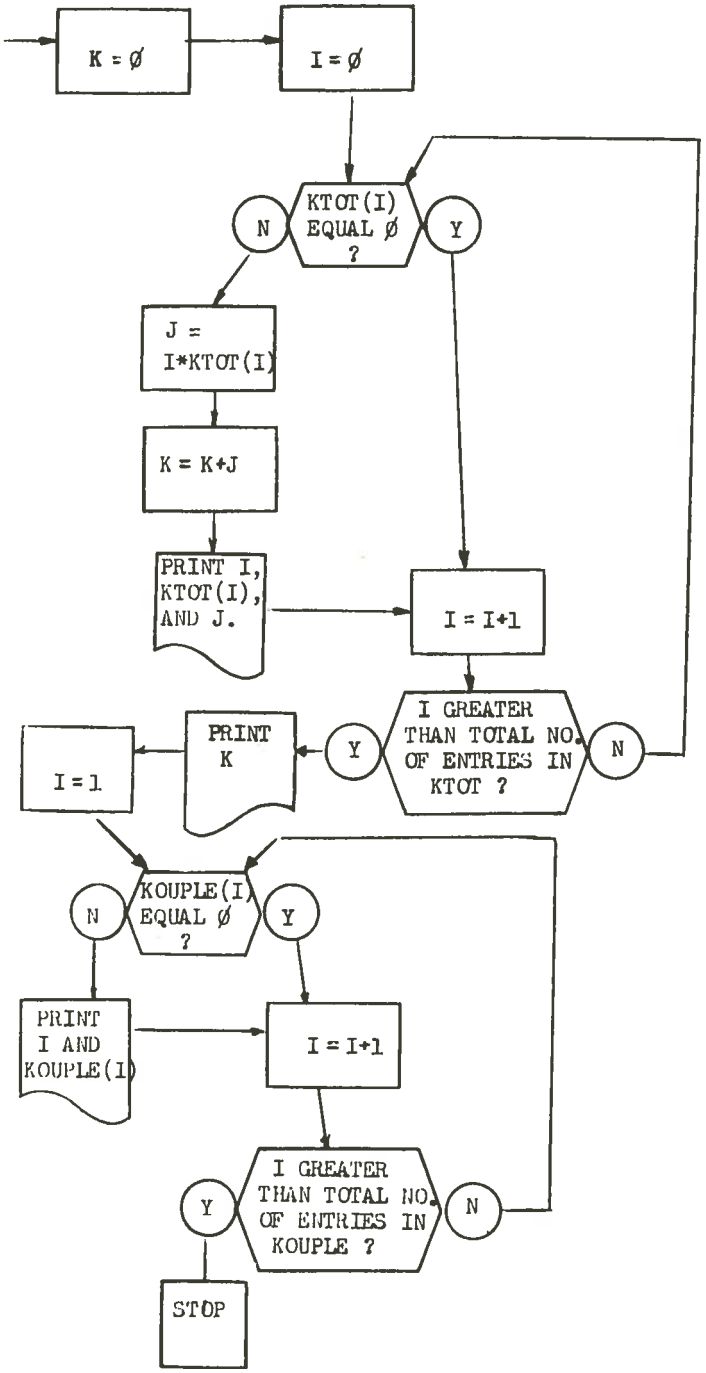
The next section of this report is devoted to a technical description of the MAPTEXT program; the description includes a flow chart, verbal outline of the program, and a complete program listing. Sally Y. Sedelow and Terry L. Ruggles are jointly responsible for this section of the document.

A. Flowchart



* SUBROUTINE SPOUT IS USED TO INSERT CHARACTERS INTO PRINT LINE AND PRINT OUT THE LINE WHEN IT IS FULL OR END OF DATA UNIT HAS BEEN REACHED.
** SET I EQUAL TO THE TOTAL NUMBER OF OCCURRENCES OF SPECIAL LINGUISTIC UNITS IN THE DATA UNIT.


OUTPUT "THEEND" TO TAPE E
REWIND E
KENT = NUNITS/60
N
REMAINDER?
Y
KENT = KENT+1
LARGEST KENT*LARGEST
KHAR = LARGEST/112
Y
REMAINDER?
N
KHAR = KHAR+1
PRINT KHAR
J = ∅
K = 1
L = ∅
READ INTO I AN ENTRY FROM TAPE E
Y
I EQUAL "THEEND"?
N
L = L+I
K = K+1
Y
K GREATER THAN KENT?
N
M = L/KHAR
Y
M LESS THAN 1?
N
N = 1
ADD AN ASTERISK TO LINE (SPOUT)
Y
N EQUAL M?
N
N = N+1
J = J+K-1
LOAD J INTO LINE
PRINT LINE
N
I EQUAL "THEEND"?
Y
PAGE 89


K = Ø
I = Ø
KTOT(I) EQUAL Ø ?
N
Y
J = I*KTOT(I)
K = K+J
PRINT I, KTOT(I), AND J.
I = I+1
I GREATER THAN TOTAL NO. OF ENTRIES IN KTOT ?
N
Y
PRINT K
I = 1
KOUPLE(I) EQUAL Ø ?
N
Y
PRINT I AND KOUPLE(I)
I = I+1
I GREATER THAN TOTAL NO. OF ENTRIES IN KOUPLE ?
Y
N
STOP

B. Description

1. Tables

Table KOUPLE (100 computer words): This table keeps track of sequential occurrences of the "special" linguistic units being represented by MAPTEXT. For example, MAPTEXT may be searching for the words MILITARY and STRATEGY in order to "map" their occurrences onto the complete textual context. The number of times that these words occur in sequence will be shown in the output from Table KOUPLE. After the entire text has been examined, entry #2 in KOUPLE will contain a number indicating how many times there are two occurrences in sequence of special linguistic units. (Entry #1 indicates the number of times special linguistic units occur alone, entry #3 indicates the number of times when there are three occurrences in sequence of special linguistic units, etc.) In our example, if MILITARY and STRATEGY were the only units specifically being sought, the value of entry #2 in KOUPLE would doubtless indicate the number of times the phrase MILITARY STRATEGY occurs in the text. A glance at the complete MAPTEXT representation would quickly confirm or negate that premise. KOUPLE is large enough to record the rather unlikely event of 100 occurrences of special characters in sequence within a data unit.

Table KTOT (1000 computer words): This table keeps track of the number of data units in which there is one occurrence of a special linguistic unit, in which there are two occurrences, and so on up through 1000 occurrences of special linguistic units per data unit. After the text has been examined,

entry #1 in KTOT contains a number indicating the number of data units in which there was just one occurrence of a special linguistic unit, entry #2 contains a number indicating the number of data units containing two occurrences, and so on, up through 1000.

Table KGROUP (100 computer words): This table is set up on the same principle as KOUPLE, except that it is used for temporary storage to keep track of sequential occurrences within a single data unit. When the end of a data unit (either sentence, paragraph, or chapter) has been reached, the values in KGROUP are stored on Tape E, so that a record for each data unit in the text will be available; at the end of each data unit, the values in KGROUP are also used to update tables KOUPLE and KTOT. For example, if entry #2 in KGROUP had a value of 4, indicating that two occurrences of special characters in sequence have appeared four times in the given data unit, entry #2 in KOUPLE will be incremented by 4. To update KTOT, the values in KGROUP are multiplied and summed so that the entry in KTOT corresponding to the number of occurrences in the given data unit can be incremented by 1. (In our example, the value of 4 in entry #2 would be multiplied by two in order to provide the value from that entry, which would then be added to the value from other entries to arrive at the sum of occurrences of special characters occuring in the given data unit.)

Tables KWORD1, KWORD2 and KSPECH (each 256 computer words): These tables contain, respectively, the first eight letters (or portion thereof if the linguistic unit has fewer than eight alphanumeric characters) of the

linguistic unit being specially sought, the second eight letters, or portion thereof, of that linguistic unit, and the character associated with that word which will represent the word in the MAPTEXT representation (e.g., MILITARY might be represented by M).

Table LINE (15 computer words): This table contains one line of output for the printer. Each computer word will hold 8 characters, and a print line has a maximum capacity of 120 characters.

Table KARD (10 computer words): This table is used to input data from punched cards. Since each computer word will hold 8 characters, 10 computer words are required for the 80 characters each punched card can contain.

Table KEAD (10 computer words): This table contains the input from one card to be used as a comment heading the output listings.

Table KTEXE (128 computer words): This table is used to store the record of occurrences of special linguistic units in each data unit in the text being examined. Table KTEXE is a temporary storage area from which information is transferred out to Tape E. A tape block on the Philco 2000 comprises 128 computer words.

Table KTEXD (128 computer words): This table is a buffer table used to input textual words from tape.

2. Items

Items I,J,K,L,M,N: These items are all used for temporary storage for various purposes throughout the program.

Item NOND: This item is used in connection with the buffer table KTEXD; the item indexes the table and indicates when a new block should be read in (during input).

Items KTEXT1, KTEXT2, KTEXT3, KTEXT4: These items are used for the temporary storage of one textual word entry. The first two items contain the textual word, and the other two contain indexing and MATCNT data used in VIA but not in MAPTEXT.

Item KEND: This item contains the Hollerith constant, THEEND, which is also stored at the end of the textual data. Thus, THEEND signals that the end of the textual data has been reached.

Item INENT: This item contains the total number of special linguistic units to be sought.

Item LEMP: This item is used for temporary storage.

Item KASE: This item contains the character used in the MAPTEXT representation for all linguistic units other than those being specially sought and represented.

Item LINK: This item keeps track of the number of special linguistic units occurring in sequence.

Item LUNIT: This item contains the character indicating whether the data unit in the text currently being examined will be defined as a sentence, paragraph, or chapter. If a sentence, LUNIT will contain one period, if a paragraph, LUNIT will contain two periods, and if a chapter, three periods.

Item NUNITS: At the end of the mapping program's operation, this item contains the value, to be used in the condensed graph, indicating the total

number of data units in the text being examined.

Item LARGEST: At the end of the mapping program's operation, this item contains the value, also to be used in the condensed graph, indicating the largest number of occurrences of special characters to have appeared in any data unit.

Item LEFT: This item holds the remainder from the arithmetic division operations occurring after statements 52 and 55 in the main program. The reasons for the division will be discussed in the description of the main program.

Item KHAR: This item contains the value indicating how many occurrences of special linguistic units per data unit will be represented by each character (an asterisk in our program) in the condensed printout, or graph, of the occurrences of those special linguistic units.

Item NONE: This item is analogous to NOND except that in this instance KTEXE and Tape E are the buffer table and tape in question and are used for both input and output.

Item KASTER: This item contains the asterisk used in the condensed printout, or graph, of the occurrences of special linguistic units.

Item NGROUP: This item contains the value indicating the size (number of entries) in Tables KOUPLE and KGROUP. In our current program, NGROUP contains the value 100.

Item NTOT: This item is analogous to NGROUP, except that NTOT indicates the size of Table KTOT.

Item KENT: This item contains the value indicating the number of data units to be compressed into a print line in the graph output.

3. Main Routine

INITIALIZE TABLES, VARIABLES AND CONSTANTS: This first section of the main program sets up constants (e.g., placing THEEND in KEND), provides initial values for variable items such as LARGEST and LINK, fills the print table, LINE, with blanks, fills tables KGROUP, KOUPLE and KTOT with zeros, and tends to other similar "housekeeping" chores. Setting up Hollerith constants (such as placing THEEND in KEND) entails using a function, LET. Although this function is described in another document (TM-1908/009/00), the description will be repeated to facilitate understanding of this program. The coding in Function LET is as follows:

```
FUNCTION LET(X)
CONTINUE
STARTTAC
          TDXL    ,0$
          TMA     ,0$
          JMP     2SRS$
          ENDTAC
END
```

The FORTRAN function statement generates code which places the address of the storage cell holding THEEND in the D-register. The address is then placed in index register 0; the instruction TMA ,0 brings the contents of that address into the A-register. The contents of the address, THEEND, are then placed in KEND via the return through 2SRS. The conventional exit from a FORTRAN subroutine is RETURN, which generates the following code:

```
TMA  N1$
JMP  3A$
```

Since THEEND exists in the A-register, the TMA N1$ instruction would destroy THEEND. In order to get the Hollerith constant safely into KEND, therefore, the RETURN statement is not used. The jump to 2SRS returns control to the main routine and is equivalent to the second instruction, JMP 3A$, generated by the RETURN statement.

The read-in of data into KARD immediately following statement 3 sets up two more Hollerith constants: in KASE, the characters to be used in the MAPTEXT representation for all "nonspecial" linguistic units, and in LUNIT, the data unit indicator (cf. explanation of Item LUNIT). The next read-in of data sets up a comment in KEAD to be used for a heading on the output which specifies the portion of the text being shown in the MAPTEXT representation.

READ IN LINGUISTIC UNITS TO BE SEARCHED FOR: This section of the program reads in the special linguistic units which the current run of MAPTEXT will study and represent with special symbols (stored, in the case of each linguistic unit, in KSPECH). For example, if the word MILITARY were a special linguistic unit, the characters M I L I T A R Y would be stored in KWORD1, KWORD2 would be blank, and the corresponding entry in KSPECH might contain the character M.

SORT SPECIAL WORDS INTO ALPHABETICAL ORDER: This section of the program uses a shuttle sort to arrange the special words in alphabetical order.

READ IN TEXT WORD OR PUNCTUATION: Subroutine IND places one word from the input text in the temporary storage slots, KTEXT1 through KTEXT4. When

KTEXT1 equals KEND, the end of the text has been reached, and the program branches to the summaries of data which produce the condensed graph and total tables.

SEARCH SPECIAL LINGUISTIC UNIT LIST FOR A MATCH: The text word which has just been brought in is first compared with the entries in KWORD1 by means of a binary search. If a match is found, the last portion of the text word is compared with entries in KWORD2. If a match is not found, the program branches to the next section of the program.

NO MATCH (OUTPUT COMMON CHARACTER): The text word or punctuation mark being examined was not one of the specially sought linguistic units; therefore, its place in the print line is filled with whatever character KASE contains--the common character. If LINK, which keeps track of the number of sequential special character occurrences, is equal to zero, the textual element which has just been examined did not mark the end of such a sequence; in this case the next textual element is simply read in and the comparison with special linguistic units begins again. If LINK is greater than zero, the textual element just examined did mark the end of a sequence of special character occurrences; in this case, the appropriate entry in KGROUP (which keeps track of sequences within a data unit) must be incremented by 1.

MATCH FOUND (OUTPUT SPECIAL CHARACTER): The text word or punctuation mark being examined was one of the specially sought linguistic units. The special character associated with that unit is placed in the temporary storage item, LEMP, and subroutine SPOUT places that special character in the appropriate spot in the print line.

CHECK TO SEE IF END OF DATA UNIT: The first three instructions in this section of the program check to see whether the data unit specified for this run of MAPTEXT is a sentence, paragraph, or chapter. If the data unit is a sentence (one period), the program branches to statement 32 to see whether the last linguistic element examined (that element is stored in LEMP) signaled the end of a sentence. Since a paragraph marker (two periods) and a chapter marker (three periods) also signal the end of a sentence, LEMP must be checked against these markers as well as against the sentence marker (one period). If the specified data unit is a paragraph, then the program branches to statement 34 to check for a chapter marker, which also signals the end of a paragraph. If the end of a data unit has not been reached, the special character in LEMP indicates that one of the linguistic units for which MAPTEXT is searching has been found; therefore, LINK, which keeps track of sequential occurrences for those special units, is incremented by 1 and the program branches back to statement 15 to read in the next textual word or punctuation mark. If the end of a data unit has been reached, the program branches to:

END OF DATA UNIT: If the end of a data unit has been reached and if LINK equals zero, there were no sequential occurrences of specially sought linguistic units in the data unit; therefore, the coding updating KGROUP and resetting LINK may be skipped. NUNITS (number of units), however, must be incremented by 1 and I must be set to zero because both will be used for the computation in:

SUM THE GROUPS IN THE DATA UNIT: This coding, which includes a DO-Loop ending at statement 44, computes the number of special linguistic unit occurrences in the data unit just ended. Each entry in KGROUP stands for that entry's

number of sequential occurrences of special linguistic units (see the description of Table KGROUP); therefore, to multiply J, which will correspond to each entry number in turn (J is initialized at 1), times the value contained in each entry, and then sum those values in I, will provide the number of occurrences in the given data unit. Subroutine OUTE then places that value in the buffer table KTEXE, for eventual transfer to Tape E.

If the number of occurrences in this data unit is larger than that in any data unit examined thus far, the item LARGEST is set to the new value. If the number of occurrences in this data unit is not larger than that in any data unit examined thus far, the program branches to:

TOTAL PER DATA UNIT: This coding increments the appropriate entry in KTOT by 1.

TOTAL PER SEQUENCE GROUP: The coding containing the DO-Loop ending at statement 48 updates KOUPLE (which keeps track of sequential occurrences for the entire text being examined) by examining each entry in KGROUP, which keeps track of sequential occurrences for each data unit in turn. As KOUPLE is updated, KGROUP is set back to zero. When this procedure has been completed, the program branches back to the read-in of the next textual word or punctuation.

FILL BUFFER WITH "THEEND": When the entire input text has been examined, the rest of the buffer table KTEXE is filled with THEENDs. This is simply to insure output to Tape E and serve as a precaution against error.

PRINT A GRAPHICAL REPRESENTATION OF SPECIAL LINGUISTIC UNIT OCCURRENCES: The program code down through statement 66 outputs the condensed version

(one page of output) of the occurrences of special linguistic units in the entire text just examined. NONE is set to 129 to signal that subroutine to read in information from tape. The first coding in TAC computes how many data units must be represented in each print line. There are 60 print lines to a page of output, so all the data units in the text must be represented in 60 lines. Hence the value in NUNITS, representing the total number of data units in the text, is divided by 60. The item KENT contains the integer value of that quotient, and the item LEFT contains the remainder. If there is no remainder, the program branches to statement 55. If there is a remainder, the value in KENT must be incremented by 1; otherwise, each line of output would end with some fraction of a data unit. Thus, at the end of this coding, KENT contains the value indicating the number of data units to be represented in each print line.

SET KHAR EQUAL TO THE NUMBER OF SPECIAL LINGUISTIC UNITS TO BE REPRESENTED BY EACH ASTERISK: The multiplication at statement 55 computes the maximum number of special linguistic units to be represented in any print line (LARGEST contains the maximum number of units in any data unit occurring in the text being examined and KENT indicates the number of those units to be represented in each print line) and stores that value in LARGEST. In the TAC coding, LARGEST is divided by 112 (the number of spaces in the print line to be used for the graph) to compute the number of special linguistic units each asterisk will represent. As in the earlier computation, the integer and fractional values are both retained, this time in KHAR and LEFT,

respectively. If there is a fractional remainder, KHAR is incremented by 1 to avoid representing one and one-half words, for example, by each asterisk. Thus, at the end of this coding, KHAR equals the number of special linguistic units to be represented by an asterisk.

START GRAPHING: The coding down to statement 71 prints out the condensed representation, or graph, of special linguistic unit occurrences. The DO-Loop ending at statement 64 reads in the record of special linguistic unit occurrences for each data unit that was saved on Tape E; each complete DO-Loop reads in the record of exactly the number of data units to be represented in each print line. When the program either branches out or falls out of the loop, L contains the number of occurrences of special linguistic units to be printed in the line representing the records just brought in during the course of the DO-Loop. The computation at statement 65 determines how many asterisks will be printed out for the data units just examined. The quotient of L divided by KHAR provides this value. If the quotient should be less than 1, nothing is printed out for that line.

OUTPUT A NUMBER OF ASTERISKS EQUIVALENT TO THE SUM TOTAL OF SPECIAL LINGUISTIC UNITS IN KENT DIVIDED BY KHAR: The DO-Loop ending with statement 66 prints out the number of asterisks specified by the value in M. The computation at statement 68 provides the value to be used in the graph printout indicating the number of data units in each print line; the value is incremented sequentially so that the location in the input text of the special linguistic units in any one print line can easily be found (e.g., 4, 8, 12, etc., up through the total number of data units in the text being examined).

The DO-Loop ending at 64 resulted in K having one more than the proper value (in order to fall out of the loop, the value in K must exceed the value in KENT--since the increments in that loop are of 1, the excess value in K will be 1), hence the subtraction of 1. The TAC coding converts the value from binary to binary-coded-decimal for use in the output. Next, Subroutine PLINE prints the line. If the end of the statistical total representing the text has not been reached, the program branches back to statement 62 and the next line is graphed. If the graphing has been completed for the entire text, the program branches to:

PRINT KTOT: The coding down to statement 74 prints out the headings and contents of Table KTOT (see description of KTOT under Tables).

PRINT KOUPLE: The coding down to the STOP statement (signaling the end of the program) prints out the heading and contents of Table KOUPLE (see description of KOUPLE under Tables).

4. Subroutines

SUBROUTINE IND (N1, N2, N3, N4): This subroutine reads textual entries stored on Tape D into Table KTEXD. Each read-in operation brings in a tape block comprising 128 computer words, or 32 textual words (each textual word entry requires 4 computer words).

SUBROUTINE INE (N1): This subroutine reads in (from Tape E to Table KTEXE) the values representing the frequency of occurrences of special linguistic units for each data unit. These values are used in the condensed representation or graph.

SUBROUTINE OUTE (N1): This subroutine reads out (from Table KTEXE to Tape E) the values representing the frequency of occurrences of special linguistic units for each data unit.

SUBROUTINE PLINE: This subroutine prints out a line (located in Table LINE) of output.

SUBROUTINE SPOUT: As the comment suggests, this subroutine places the special characters representing linguistic units in the print line until the line is filled, or until no more data are required for the line; at this point, the line is printed. Before calling SPOUT, the main routine places the appropriate character in the call statement as a parameter. This character is contained in N1 at the beginning of the SPOUT subroutine. SPOUT first checks to see if the special character is one indicating the end of a data unit. If the end of a sentence is signaled by either a period, question mark, or exclamation point, SPOUT branches to EOS and prints out one period. If the end of a paragraph is signaled (by two periods), SPOUT branches to EOP and prints out two periods. If the end of a chapter is signaled (by three periods), SPOUT branches to EOC and prints out three periods. If the special character does not indicate the end of a data unit, SPOUT branches to SP, where the special character is loaded into the print line. In general, the coding under SP uses masks to position the special character correctly in the print line. The comparison of 119 and the value in POS (which keeps track of the position in which the character is to be placed) shows when the print line has been filled. When it has been filled, Subroutine PLINE prints the line. If the line is not filled, the program

branches to S1 to compute the proper print position for the character. For example, if the Q-register contains the value 57, the division of that value by the dividend, 8, will result in a quotient of 7 with a remainder of 1. Remembering that eight characters will fit into any one word in the 15-word print line, this quotient means that the character will occupy the first position in the eighth word (LINE(8)) of the print line. This information enables the program to select the proper mask for insertion of the special character into entry #8 in Table LINE.

C. MAPTEXT Program Listing

```
                    IDENTIFY F, 16K, 8X, 2500W      I        MAPTEXT
C                                                   READ IN A TEXT ENTRY
      SUBROUTINE IND(N1,N2,N3,N4)
      COMMON I,J,K,L,M,N
      COMMON KTEXD,NOND
      COMMON KTEXT1,KTEXT2,KTEXT3,KTEXT4
      COMMON KEND,INENT,LEMP,KARD,LINE,KASE
      COMMON KGROUP,KOUPLE,KTOT,KTEXE,KEAD
      COMMON LINK,LUNIT,NUNITS,LARGEST,LEFT
      COMMON KHAR,NONE,KASTER,NGROUP,NTOT
      COMMON KWORD1,KWORD2,KSPECH,KENT
      DIMENSION  KGROUP(100)
      DIMENSION  KOUPLE(100)
      DIMENSION  KTOT(1000)
      DIMENSION  KTEXD(128)
      DIMENSION  KTEXE(128)
      DIMENSION KEAD(10)
      DIMENSION  KARD(10)
      DIMENSION LINE(15)
      DIMENSION  KWORD1(256)
      DIMENSION  KWORD2(256)
      DIMENSION  KSPECH(256)
      IF (NOND)GTE(128),GO TO 20
   10 N1 = KTEXD(NOND)
      N2 = KTEXD(NOND+1)
      N3 = KTEXD(NOND+2)
      N4 = KTEXD(NOND+3)
      NOND = NOND+4
      RETURN
   20 STARTTAC
        R       JMP     IOINT$
                N/6T23;N/1T39$
                HLT     TZZD $
                HLT     KTEXD$
                ENDTAC
      NOND = 1
      GO TO 10
      END
C                                                   READ IN FREQUENCY COUNT
C                                                   COMPUTED FROM KGROUP
      SUBROUTINE INE(N1)
      COMMON I,J,K,L,M,N
      COMMON KTEXD,NOND
      COMMON KTEXT1,KTEXT2,KTEXT3,KTEXT4
      COMMON KEND,INENT,LEMP,KARD,LINE,KASE
      COMMON KGROUP,KOUPLE,KTOT,KTEXE,KEAD
      COMMON LINK,LUNIT,NUNITS,LARGEST,LEFT
      COMMON KHAR,NONE,KASTER,NGROUP,NTOT
      COMMON KWORD1,KWORD2,KSPECH,KENT
      DIMENSION  KGROUP(100)
      DIMENSION  KOUPLE(100)
      DIMENSION  KTOT(1000)
      DIMENSION  KTEXD(128)
      DIMENSION  KTEXE(128)
      DIMENSION KEAD(10)
      DIMENSION  KARD(10)
      DIMENSION LINE(15)
      DIMENSION  KWORD1(256)
      DIMENSION  KWORD2(256)
```

```
                                                        I       MAPIEXT
      DIMENSION  KSPECH(256)
      IF (NONE)GT(128),GO TO 20
   10 N1 = KTEXE(NONE)
      NONE = NONE+1
      RETURN
   20 STARTTAC
        R       JMP     IOINT $
                N/4T23;N/1T39 $
                HLI     TZZE  $
                HLI     KTEXE $
                ENDTAC
      NONE= 1
      GO TO 10
      END
C                                          READ OUT FREQUENCY COUNT
C                                          COMPUTED FROM KGROUP
      SUBROUTINE OUTE(N1)
      COMMON I,J,K,L,M,N
      COMMON KTEXD,NOND
      COMMON KTEXT1,KTEXT2,KTEXT3,KTEXT4
      COMMON KEND,INENT,LEMP,KARD,LINE,KASE
      COMMON KGROUP,KOUPLE,KTOT,KTEXE,KEAD
      COMMON LINK,LUNIT,NUNITS,LARGEST,LEFT
      COMMON KHAR,NONE,KASTER,NGROUP,NTOT
      COMMON KWORD1,KWORD2,KSPECH,KENT
      DIMENSION  KGROUP(100)
      DIMENSION  KOUPLE(100)
      DIMENSION  KTOT(1000)
      DIMENSION  KTEXD(128)
      DIMENSION  KTEXE(128)
      DIMENSION KEAD(10)
      DIMENSION  KARD(10)
      DIMENSION LINE(15)
      DIMENSION  KWORD1(256)
      DIMENSION  KWORD2(256)
      DIMENSION  KSPECH(256)
      KTEXE(NONE) = N1
      NONE = NONE+1
      IF (NONE)GT(128),GO TO 10
      RETURN
   10 STARTTAC
        R       JMP     IOINT $
                N/5T23;N/1T39 $
                HLI     TZZE  $
                HLI     KTEXE $
                ENDTAC
      NONE = 1
      RETURN
      END
C                                          PRINTS LINE
      SUBROUTINE PLINE
      COMMON I,J,K,L,M,N
      COMMON KTEXD,NOND
      COMMON KTEXT1,KTEXT2,KTEXT3,KTEXT4
      COMMON KEND,INENT,LEMP,KARD,LINE,KASE
      COMMON KGROUP,KOUPLE,KTOT,KTEXE,KEAD
      COMMON LINK,LUNIT,NUNITS,LARGEST,LEFT
      COMMON KHAR,NONE,KASTER,NGROUP,NTOT
      COMMON KWORD1,KWORD2,KSPECH,KENT
```

```
                                                        I       MAPTEXT
      DIMENSION  KGROUP(100)
      DIMENSION  KOUPLE(100)
      DIMENSION  KTOT(1000)
      DIMENSION  KTEXD(128)
      DIMENSION  KTEXE(128)
      DIMENSION KEAD(10)
      DIMENSION  KARD(10)
      DIMENSION LINE(15)
      DIMENSION  KWORD1(256)
      DIMENSION  KWORD2(256)
      DIMENSION  KSPECH(256)
    1 FORMAT (1H 15A8)
      PRINT 1,LINE
      STARTTAC
                TMD     L/LINE$
                TDXLC   ,1$
                TMD     W/          $
        L       RPTA    15$
                TDM     1,1$
                ENDTAC
      RETURN
      END
C                                       PUTS THE FIRST CHARACTER OF
C                                       N1 INTO SUCCESSIVE POSITIONS
C                                       OF THE PRINT LINE
C                                       AND PRINTS LINE WHEN FULL
C                                       OR WHEN NO MORE
C                                       DATA FOR THE LINE
      SUBROUTINE SPOUT(N1)
      COMMON I,J,K,L,M,N
      COMMON KTEXD,NOND
      COMMON KTEXT1,KTEXT2,KTEXT3,KTEXT4
      COMMON KEND,INENT,LEMP,KARD,LINE,KASE
      COMMON KGROUP,KOUPLE,KTOT,KTEXE,KEAD
      COMMON LINK,LUNIT,NUNITS,LARGEST,LEFT
      COMMON KHAR,NONE,KASTER,NGROUP,NTOT
      COMMON KWORD1,KWORD2,KSPECH,KENT
      DIMENSION  KGROUP(100)
      DIMENSION  KOUPLE(100)
      DIMENSION  KTOT(1000)
      DIMENSION  KTEXD(128)
      DIMENSION  KTEXE(128)
      DIMENSION KEAD(10)
      DIMENSION  KARD(10)
      DIMENSION LINE(15)
      DIMENSION  KWORD1(256)
      DIMENSION  KWORD2(256)
      DIMENSION  KSPECH(256)
      CONTINUE
      STARTTAC
                TMA     N1    $
                TMD     W/.         $
                JAED    EOS   $           END OF SENTENCE
                TMD     W/..        $
                JAED    EOP   $           END OF PARAGRAPH
                TMD     W/...       $
                JAED    EOC   $           END OF CHAPTER
                TMD     W/?         $
                JAED    EOS   $
```

```
                                                        MAPTEXT
         TMD     W/'.       $EXCLAMATION POINT
         JAED    EOS   $
         JMP     SP    $
         TMA     N1    $
         JMP     2SRS  $
SP       TJM     EXIT  $         LOAD SPECIAL CHARACTER
         TMD     L/MASK$         INTO PRINT LINE
         TDXLC   ,4    $
         TMD     L/LINE$
         TDXLC   ,5    $
         TMQ     POS   $         POSITION INTO WHICH THE
         TMA     D/119$          CHARACTER IS TO BE PLACED
         JAGQ    S1    $
S        PLINE         $
         TMQ     D/8   $
         TQM     POS   $
S1       CA            $         POS/8 = WORD.CHARACTER
         DAQ     D/8   $
         SLAQ    8     $
         TQD           $
         ADXR    ,5    $     LINE (WORD)
         TAD           $
         ADXR    ,4    $     MASK (CHAR)
         TMQ     D/6   $
         MA            $
         TQD           $
         TDXRC   ,3    $     SHIFT (WORD)
         TMA     N1    $
         SRA     ,3    $
         TMQ     ,4    $
         EIS     ,5    $
         TMA     D/1   $
         AMS     POS   $         BUMP POS
EXIT     JMP     (P)   $
EOS      JMP     SP    $         OUTPUT ONE PERIOD
         JMP     EOZ   $
EOP      JMP     SP    $         OUTPUT TWO PERIODS
         JMP     SP    $
         JMP     EOZ   $
EOC      JMP     SP    $         OUTPUT THREE PERIODS
         JMP     SP    $
         JMP     SP    $
         JMP     EOZ   $
POS      D/8           $
MASK     0/0077777777777777$     FOR EXTRACT AND INSERT
         0/7700777777777777$     OF A SPECIAL CHARACTER
         0/7777007777777777$     INTO ONE OF THE EIGHT
         0/7777770077777777$     POSSIBLE CHARACTER POSITIONS
         0/7777777700777777$     OF A COMPUTER WORD
         0/7777777777007777$
         0/7777777777770077$
         0/7777777777777700$
SENT     D/1           $
EOZ      TMD     D/8   $
         TDM     POS   $
S        PLINE   $
         TMA     D/1   $
         AMS     SENT  $
S        BIN2BCD $
```

```
                                                          I       MAPTEXT
      S       ZSUPP   $
              TMQ     0/6060$
              SLAQ    12   $
              TAM     LINE $
              ENDTAC
      RETURN
      END
C                                                  OUTPUTS A HOLLERITH CONSTANT
      FUNCTION LET(X)
      CONTINUE
      STARTTAC
              TDXL    ,0$
              TNA     ,0$
              JMP     2SRS$
              ENDTAC
      END
      COMMON I,J,K,L,M,N
      COMMON KTEXD,NCND
      COMMON KTEXT1,KTEXT2,KTEXT3,KTEXT4
      COMMON KEND,INENT,LEMP,KARD,LINE,KASE
      COMMON KGROUP,KOUPLE,KTOT,KTEXE,KEAD
      COMMON LINK,LUNIT,NUNITS,LARGEST,LEFT
      COMMON KHAR,NONE,KASTER,NGROUP,NTOT
      COMMON KWORD1,KWORD2,KSPECH,KENT
      DIMENSION  KGROUP(100)
      DIMENSION  KOUPLE(100)
      DIMENSION  KTOT(1000)
      DIMENSION  KTEXD(128)
      DIMENSION  KTEXE(128)
      DIMENSION KEAD(10)
      DIMENSION  KARD(10)
      DIMENSION LINE(15)
      DIMENSION  KWORD1(256)
      DIMENSION  KWORD2(256)
      DIMENSION  KSPECH(256)
C                                                  INITIALIZE TABLES,
C                                                  VARIABLES AND CONSTANTS
      KASTER = LET(8H*       )
      KEND = LET(8HTHEEND  )
      LINE(1) = LET(8H     1  )
      DO 1 I = 2,15,1
      LINE(I) = LET(8H        )
    1 CONTINUE
      LINK = 0
      NUNITS = 0
      LARGEST = 0
      NONE = 1
      NGROUP = 100
      NTOT = 1000
      DO 2 I = 1,NGROUP,1
      KGROUP(I) = 0
      KOUPLE(I) = 0
    2 CONTINUE
      DO 3 I = 1,NTOT,1
      KTOT(I) = 0
    3 CONTINUE
      READ 103,KARD
      KASE = KARD(2)
      LUNIT = KARD(3)
```

```
                                                        I       MAPTEXT
      READ 103,KEAD
      PRINT 100,KEAD
      PRINT 101
      I = 1
C                                          READ IN LINGUISTIC
C                                          UNITS TO BE
C                                          SEARCHED FOR
   4  READ 103,KARD
      IF (KARD(1))E(KEND),GO TO 5
      KWORD1(I) = KARD(1)
      KWORD2(I) = KARD(2)
      KSPECH(I) = KARD(3)
      I = I+1
      GO TO 4
C                                          SORT SPECIAL
C                                          LINGUISTIC UNITS INTO
C                                          ALPHABETICAL ORDER
   5  INENT = I-1
      REWIND 3
   6  J = 2
   7  I = J-1
      IF (KWORD1(I))E(KWORD1(J)),GO TO 13
   8  IF (KWORD1(J))GT(KWORD1(I)),GO TO 11
   9  LEMP = KWORD1(J)
      KWORD1(J) = KWORD1(I)
      KWORD1(I) = LEMP
      LEMP = KWORD2(J)
      KWORD2(J) = KWORD2(I)
      KWORD2(I) = LEMP
      LEMP = KSPECH(J)
      KSPECH(J) = KSPECH(I)
      KSPECH(I) = LEMP
      J = J-1
      IF (J)LT(2),GO TO 6
      GO TO 7
  11  J = J+1
      IF (J)GT(INENT),GO TO 14
      GO TO 7
  13  IF (KWORD2(J))GTE(KWORD2(I)),GO TO 11
      GO TO 9
  14  DO 145 I = 1,INENT,1
      PRINT 102,KWORD1(I),KWORD2(I),KSPECH(I)
 145  CONTINUE
      PRINT 104
      PRINT 100,KEAD
      NOND = 128
C                                          READ IN TEXT WORD
C                                          OR PUNCTUATION
  15  CALL IND(KTEXT1,KTEXT2,KTEXT3,KTEXT4)
      IF (KTEXT1)E(KEND),GO TO 50
C                                          SEARCH SPECIAL
C                                          LINGUISTIC UNIT LIST
C                                          FOR A MATCH
      J = 0
      K = INENT+1
  16  I = (J+K)/2
      IF (I)E(K),GO TO 20
      IF (I)E(J),GO TO 20
      IF (KTEXT1)E(KWORD1(I)),GO TO 19
```

```
                                                          I        MAPTEXT
      IF (KTEXT1)GI(KWORD1(I)),GO TO 18
   17 K = I
      GO TO 16
   18 J = I
      GO TO 16
   19 IF (KTEXT2)E(KWORD2(I)),GO TO 30
      IF (KTEXT2)GI(KWORD2(I)),GO TO 18
      GO TO 17
C                                          NO MATCH
C                                          OUTPUT COMMON CHARACTER
   20 CALL SPOUT(KASE)
      IF (LINK)E(0),GO TO 15
      KGROUP(LINK) = KGROUP(LINK)+1
      LINK = 0
      GO TO 15
C                                          MATCH FOUND
C                                          OUTPUT SPECIAL CHARACTER
   30 LEMP = KSPECH(I)
      CALL SPOUT(LEMP)
C                                          CHECK TO SEE IF END OF
C                                          DATA UNIT
      IF (LUNIT)E(LET(8H.        )),GO TO 32
      IF (LUNIT)E(LET(8H..       )),GO TO 34
      IF (LUNIT)E(LET(8H...      )),GO TO 36
   32 IF (LEMP)E(LET(8H..       )),GO TO 40
   34 IF (LEMP)E(LET(8H...      )),GO TO 40
   36 IF (LUNIT)E(LEMP),GO TO 40
      LINK = LINK+1
      GO TO 15
C                                          END OF DATA UNIT
   40 IF (LINK)E(0),GO TO 42
      KGROUP(LINK) = KGROUP(LINK)+1
      LINK = 0
   42 NUNITS = NUNITS+1
      I = 0
C                                          SUM THE GROUPS IN
C                                          THE DATA UNIT
      DO 44 J = 1,NGROUP,1
      I = I+J*KGROUP(J)
   44 CONTINUE
      CALL OUTE(I)
      IF (I)LTE(LARGEST),GO TO 46
      LARGEST = I
C                                          TOTAL PER DATA UNIT
   46 KTOT(I) = KTOT(I)+1
C                                          TOTAL PER SEQUENCE GROUP
      DO 48 J = 1,NGROUP,1
      KOUPLE(J) = KOUPLE(J)+KGROUP(J)
      KGROUP(J) = 0
   48 CONTINUE
      GO TO 15
C                                          FILL BUFFER WITH (THEEND)
   50 DO 52 I = NONE,128,1
      CALL OUTE(KEND)
   52 CONTINUE
      REWIND 3
      REWIND 4
C                                          PRINT A GRAPHICAL
C                                          REPRESENTATION OF
```

```
                                                    I       MAPIEXT
C                                                   SPECIAL LINGUISTIC
C                                                   UNIT OCCURRENCES
      NONE = 129
      STARTTAC
                CA      $
                TMQ     NUNITS$
                DAQ     N/60T15$            NUNITS/60 = KENT.LEFT
                SLQ     32$
                TQM     KENT$
                TAM     LEFT$
                ENDTAC
      IF (LEFT)E(0),GO TO 55
C                                           SET KENT EQUAL
C                                           TO THE NUMBER OF
C                                           DATA UNITS TO BE COMPRESSED
C                                           INTO A PRINT LINE
      KENT = KENT+1
   55 LARGEST = KENT+LARGEST
      STARTTAC
                CA      $
                TMQ     LARGEST$
                DAQ     N/112T15$           LARGEST/112 = KHAR.LEFT
                SLQ     32$
                TQM     KHAR$
                TAM     LEFT$
                ENDTAC
      IF (LEFT)E(0),GO TO 60
C                                           SET KHAR EQUAL
C                                           TO THE NUMBER OF
C                                           SPECIAL LINGUISTIC
C                                           UNITS TO BE
C                                           REPRESENTED BY AN ASTERISK
      KHAR = KHAR+1
C                                           START GRAPHING
   60 PRINT 105,KASTER,KHAR
      J = 0
   62 L = 0
C                                           READ IN KENT-NUMBER-
C                                           OF-DATA-UNITS
      DO 64 K = 1,KENT,1
      CALL INE(I)
      IF (I)E(KEND),GO TO 65
      L = L+I
   64 CONTINUE
   65 M = L/KHAR
      IF (M)LT(1),GO TO 68
C                                           OUTPUT A NUMBER OF
C                                           ASTERISKS EQUIVALENT
C                                           TO THE SUM TOTAL OF
C                                           SPECIAL LINGUISTIC
C                                           UNITS IN KENT-NUMBER-OF
C                                           DATA-UNITS DIVIDED
C                                           BY KHAR
      DO 66 N = 1,M,1
      CALL SPOUT(KASTER)
   66 CONTINUE
   68 J = J+K-1
      STARTTAC
                TMA     J$                  CONVERTS DATA UNIT
```

```
                SRA     32$                 COUNT FOR PRINTING MAPTEXT
          S     BTN28CD $
          S     ZSUPP   $
                TMU     0/6060$
                SLAQ    12$
                TAM     LINES$
                TMU     D/8$
                TDM     SPOUT.POS$
                ENDTAC
          CALL PLINE
          IF (I)E(KEND),GO TO 71
          GO TO 62
C                                           PRINT KTOT
   71     PRINT 104
          PRINT 100,KEAD
          PRINT 106
          K = 0
          DO 74 I = 0,NTOT,1
          IF (KTOT(I))E(0),GO TO 74
          J = I*KTOT(I)
          K = K+J
          PRINT 107,I,KTOT(I),J
   74     CONTINUE
          PRINT 108,K
C                                           PRINT KOUPLE
          PRINT 109
          J = NGROUP-1
          DO 80 I = 1,J,1
          IF (KOUPLE(I))E(0),GO TO 80
          PRINT 110,I,KOUPLE(I)
   80     CONTINUE
          STOP
  100 FORMAT(1H ,10A8)
  101 FORMAT(1H0,32H   KEY TO MAPTEXT REPRESENTATION)
  102 FORMAT(1H 3A8)
  103 FORMAT (10A8)
  104 FORMAT(1H1)
  105 FORMAT(1H1,1A2,1H=,1I4,14H OCCURRENCE(S))
  106 FORMAT(1H0,35H TOTALS OF OCCURRENCE IN DATA UNITS)
  107 FORMAT(1H 3I8)
  108 FORMAT(1H ,16H         TOTAL   ,1I8)
  109 FORMAT(1H0,33H NUMBER OF SEQUENTIAL OCCURRENCES)
  110 FORMAT(1H 2I8)
          COMPLETE
          END$
```

IV. Professional Activities

1. Sedelow, S. Y. Computers in Linguistic Analysis. Presented as a public lecture sponsored by the English Department at Southern Illinois University, Carbondale, Illinois, February 19, 1965.
2. Sedelow, S. Y. Stylistic Analysis, Report on the First Year of Research. SDC document TM-1908/100/00, Santa Monica, California, System Development Corporation, March 1, 1965.
3. Sedelow, S. Y. Quarterly Report, 1 March 1965 to 1 June 1965. SDC document TM-(L)-1908/006/00, Santa Monica, California, System Development Corporation, May 25, 1965.
4. Sedelow, S. Y. Computers and Stylistics. Presented at the annual meeting of Missouri Academy of Science, St. Louis, Missouri, April 24, 1965.
5. Sedelow, S. Y. Form Recognition in Literature. Presented at Symposium on Form Recognition in Medicine, Music, Literature, and Law. Presented at the IFIPS Meeting, New York, May 28, 1965.
6. Sedelow, S. Y. Quarterly Report, 1 June 1965 to 1 September 1965. SDC document TM-(L)-1908/007/00, Santa Monica, California, System Development Corporation, August 24, 1965.
7. Sedelow, S. Y. Stylistic Analysis. Presented to the Naval Intelligence Research Advisory Group, Washington, D. C., June 29, 1965.
8. Sedelow, S. Y. Quarterly Report, 1 September 1965 to 1 December 1965. SDC document TM-(L)-1908/008/00, Santa Monica, California, System Development Corporation, November 19, 1965.

9. Sedelow, S. Y. Form Recognition in Verbal Data. Presented at the St. Louis ACM Chapter, St. Louis, Missouri, October 21, 1965.

10. Sedelow, S. Y. Computers and Research in Literature. Presented at the Computers and Research in the Humanities Conference at Purdue University, Lafayette, Indiana, October 29, 1965.

11. Sedelow, S. Y. Computers and the Study of Natural Language. Presented at the Texas A & M University, College Station, Texas, December 3, 1965.

12. Sedelow, S. Y. and Ruggles, T. L. Updating of THESAUR Program. SDC document TM-1908/009/00, Santa Monica, California, System Development Corporation, December 17, 1965.

13. Sedelow, S. Y. Chairman,Conference on the Application of Computing Devices to the Study of Language and Literature, Modern Language Association, Chicago, Illinois, December 27, 1965.

14. Sedelow, S. Y. Learning the New Methodologies. Presented at the American Educational Research Association, Chicago, Illinois, February 18, 1966.

Appendix:

Complete VIA Output, in Alphabetic Range A - D,

for RAND Translation, Sokolovsky, Chapter 1.

(VIA runs on the Praeger and RAND translations were set up so that the RAND run was considered a run on a second section of text using the thesaurus constructed during the first run. Therefore, in this output, all words preceded by dashes are words occurring in RAND for the first time, i.e., they did not occur in Praeger.)

THESAUR NC VAPS 02-28-66 MERGE 00000050 RAND CHAPTER ONE

1	ABILITY	4
2	ABLE	3
3	ABSENCE	2
4	ACCELERATED	2
5	ACCEPTANCE	2
	ACCEPTED	
6	ACCESS	1
7	ACCIDENTAL	3
8	ACCOMPLISHED	18
	ACCOMPLISHMENT	
	ACCOMPLISHMENTS	
	ACCOMPLISH	
9	ACCORDANCE	15
	ACCORDING	
	ACCORD	
10	ACCOUNT	13
	ACCOUNTS	
11	ACCUMULATED	4
	ACCUMULATION	
12	ACCURATELY	3
	ACCURATE	
13	ACHIEVED	21
	ACHIEVEMENT	
	ACHIEVE	
	ACHIEVING	
14	ACQUIRES	1
15	ACTED	14
	ACTION	
	ACTIONS	
	ACT	
	ACTUALLY	
	ACTUAL	
16	ACTIVELY	1
17	ACTIVITIES	6
	ACTIVITY	
18	ADAPTED	3
	ADAPTING	
19	ADDITIONAL	8
	ADDITION	
20	ADHERED	2
	ADHERENTS	
21	ADJUST	2
22	ADMINISTRATION	9
	ADMINISTRATIVE	
23	ADMISSION	1
24	ADMITS	1
25	ADVANCED	9
	ADVANCE	
26	ADVANTAGE	9
	ADVANTAGES	
27	ADVENT	1
28	ADVENTURISTIC	2
29	ADVOCATING	1
30	*AEGIS	1
	PROTECTION (DIFFERENT FORM APPEARS IN THE TEXT)	
31	AFFAIRS	2
32	AFFECTED	3
	AFFECT	

	AFFECTS			
33	AFFILIATED			1
34	AFRAID			1
35	AFRICAN			2
	AFRICA			
36	AGAIN			2
37	AGAINST			18
38	AGENCIES			11
	AGENCY			
39	AGES			1
40	AGGRAVATION			1
41	*AGGREGATE			1
		ALL		52
		--------AGGREGATE		
		ALTOGETHER		
		COMPLETE		
		ENTIRE		
		EVERY		
		INTEGRATE	(DIFFERENT FORM APPEARS IN THE TEXT)	
		NO		
		TOTALITY		
		UNITY		76
		ACCORD		
		AGREEMENT		
		DISAGREE	(DIFFERENT FORM APPEARS IN THE TEXT)	
		SOLIDARITY		
		WHOLE		
		SUM		
		TOTALNESS	(DIFFERENT FORM APPEARS IN THE TEXT)	
		TOTALITY		
		WHOLE		
42	AGGRESSION			18
	AGGRESSIVE			
	AGGRESSORS			
43	AGREED			8
	AGREEMENT			
	AGREEMENTS			
44	AGRICULTURAL			2
	AGRICULTURE			
45	AIDED			3
	AID			
46	AIMED			54
	AIM			
		ASPIRE		
		BEARING		
		COURSE		
		DESIGN		
		DIRECT		
		END		
		GOAL		
		INTEND	(DIFFERENT FORM APPEARS IN THE TEXT)	
		INTENT	(DIFFERENT FORM APPEARS IN THE TEXT)	
		OBJECT	(DIFFERENT FORM APPEARS IN THE TEXT)	
		OBJECTIVE		
		POINT		
		PURPOSE		
		TARGET		
	AIMS			
47	AIRCRAFT			5
48	AIRPLANES			2

```
49  AIR                                                                    7
50  A.D.                                                                   1
51  ALEXANDER                                                              1
52  ALLEGEDLY                                                              4
53  ALLIANCE                                                              31
    ALLIANCES
    ALLIED
    ALLIES
    ALLY
54  ALL-ENCOMPASSING                                                       1
55  ALL-INCLUSIVE                                                          1
56  ALL-OUT                                                                1
57 *ALLOCATE                                                               1
                    DISPOSE                          (DIFFERENT FORM APPEARS IN THE TEXT)
                    GROUP
58 *ALLOW                                                                  1
                    LET
                    PERMIT
                    PROHIBIT
59  ALL                                                                   52
    *       --------AGGREGATE                                             1
                            ALL
                            SUM
                            TOTALNESS                (DIFFERENT FORM APPEARS IN THE TEXT)
                            TOTALITY
                            WHOLE
                    ALTOGETHER
                    COMPLETE
                    ENTIRE
                    EVERY
                    INTEGRATE                        (DIFFERENT FORM APPEARS IN THE TEXT)
                    NO
                    TOTALITY
                    UNITY                                                  76
                            ACCORD
                            AGREEMENT
                            DISAGREE                 (DIFFERENT FORM APPEARS IN THE TEXT)
                            SOLIDARITY
                    WHOLE
60  ALMOST                                                                 5
61  ALONE                                                                  2
62  ALREADY                                                                6
63 *ALTERED                                                                1
                    CHANGED
    *       --------DEFLECTED                       (DIFFERENT FORM APPEARS IN THE TEXT)   1
                    --------SHIFT
                            TURN
    *                       DIVERT                  (DIFFERENT FORM APPEARS IN THE TEXT)   0
                                    (PRECEDING PRIMARY WORD DOES NOT APPEAR IN THIS SECTION)
                            --------DEFLECT
                            --------SHIFT
                                    TURN
64  ALTHOUGH                                                               4
65  ALTOGETHER                                                             1
66  ALWAYS                                                                14
67  AMERICA'S-STRATE                                                       1
68  AMERICAS                                                              31
    AMERICAN
    AMERICANS
    AMERICA
```

69	AMERICAN-KOREAN			1
70	AMMUNITION			1
71	AMOUNTS			1
72	ANALYSIS			8
73	ANARCHY			1
74	ANCIENT			3
75	ANGLO-AMERICAN			3
76	ANGLO-FRENCH			1
77	ANNIHILATE			5
	ANNIHILATION			
78	ANNUALLY			4
	*ANNUAL			
		PERIODIC	(DIFFERENT FORM APPEARS IN THE TEXT)	
		YEARLY	(DIFFERENT FORM APPEARS IN THE TEXT)	
79	ANTAGONISTIC			2
	ANTAGONISTS			
80	*ANTICIPATED			1
		EXPECTED		
		FORESEEN	(DIFFERENT FORM APPEARS IN THE TEXT)	
		PREDICTED	(DIFFERENT FORM APPEARS IN THE TEXT)	
81	ANTI-GERMAN			1
82	ANTI-HITLER			1
83	ANTI-SCIENTIFIC			2
84	ANTI-SOVIET			1
85	ANTIQUITY			2
86	ANZUS			1
87	APART			1
88	APPARATUS			2
89	APPEARANCE			12
	APPEARED			
	APPEARS			
90	APPLICATION			8
	APPLIED			
	APPLY			
91	APPRAISES			1
92	APPROACH			2
93	APPROPRIATE			1
94	APPROVED			2
	APPROVE			
95	APPROXIMATELY			4
96	APRIL			1
97	AREA			16
	AREAS			
98	ARENA			2
99	ARGUED			1
100	ARISEN			10
	ARISE			
	ARISES			
	ARISING			
101	ARMAMENT-CHANGES			1
102	ARMAMENT			146
	ARMAMENTS			
	ARMED			
	ARMIES			
	ARMS			
	ARMY'S			
	ARMY			
	(DIFFERENT FORM APPEARS IN THE THESAURUS)	ARM		
	--------AMMUNITION			
	DISARM		(DIFFERENT FORM APPEARS IN THE TEXT)	

FIREARM			
GUN			
MISSILE			
NUCLEAR			
STEEL			
WAR			321
BALLISTICS			
BATTLEFIELD			
BATTLE			
BELLIGERENT			
CAMPAIGN			
COMBAT			
FIGHTING			
GUNPOWDER			
MILITARY			292
FIGHTER		(DIFFERENT FORM APPEARS IN THE TEXT)	
GENERAL			55
--------AGGREGATE			1
	ALL		
	SUM		
	TOTALNESS	(DIFFERENT FORM APPEARS IN THE TEXT)	
	TOTALITY		
	WHOLE		
COMMON			
HABITUAL		(DIFFERENT FORM APPEARS IN THE TEXT)	
NATURAL			
NORMAL		(DIFFERENT FORM APPEARS IN THE TEXT)	
REGULAR			
TOTAL			
TYPICAL			
WHOLE			
GUNMAN		(DIFFERENT FORM APPEARS IN THE TEXT)	
MAJOR			
MARSHAL			
OFFICER			
SOLDIER			
WARFARE			
WAR			
PACIFIC			
PEACE			
STRATEGY			297
GENERALSHIP		(DIFFERENT FORM APPEARS IN THE TEXT)	55
--------AGGREGATE			1
	ALL		
	SUM		
	TOTALNESS	(DIFFERENT FORM APPEARS IN THE TEXT)	
	TOTALITY		
	WHOLE		
COMMON			
HABITUAL		(DIFFERENT FORM APPEARS IN THE TEXT)	
NATURAL			
NORMAL		(DIFFERENT FORM APPEARS IN THE TEXT)	
REGULAR			
TOTAL			
TYPICAL			
WHOLE			
PLAN			
TACTICS			
TACTICS			
CONFLICT		(DIFFERENT FORM APPEARS IN THE TEXT)	0

(PRECEDING PRIMARY WORD DOES NOT APPEAR IN THIS SECTION)

BATTLE
DISAGREEMENT
--------IRRECONCILABLE
OPPOSE
WAR

WEAPON

103 ARMISTICE
104 AROUND 1
105 ARRANGEMENTS 1
*ARRANGE 2

DESIGN
* --------DEVISE
DESIGN 2
INVENT (DIFFERENT FORM APPEARS IN THE TEXT)
OUTLINE
PLAN
ORGANIZE
PLAN
--------SYSTEMATIC

106 ARRIVAL
107 ARRIVING 1
108 ARTICLE 1
109 ARTIFICIAL 4
110 ARTILLERY 1
111 ART 6
ARTS 15
112 *ASCERTAIN

CONCLUDE 1
DETERMINE (DIFFERENT FORM APPEARS IN THE TEXT)
GATHER (DIFFERENT FORM APPEARS IN THE TEXT)
JUDGE (DIFFERENT FORM APPEARS IN THE TEXT)

113 ASK
114 ASPECT 2
ASPECTS 5
115 ASPIRATION
ASPIRATIONS 1
ASPIRES
116 ASSERTED
ASSERTS 2
117 ASSIGNED
ASSIGNMENTS 5
118 ASSISTANCE
119 ASSUMED 2
ASSUME 6
ASSUMES
ASSUMPTIONS
120 ASSURE
ASSURES 10
121 ATLANTIC
122 *ATMOSPHERE 2

--------SKY 1
SPACE

123 ATOMIC
124 ATTACKED 2
ATTACKING 13
ATTACK
ATTACKS
125 ATTAINED
ATTAINMENT 5

126	ATTEMPTED		7
	ATTEMPTING		
	ATTEMPT		
	ATTEMPTS		
127	ATTENTION		8
128	ATTITUDE		2
	ATTITUDES		
129	ATTRIBUTED		2
	ATTRIBUTES		
130	*ALGMENTED		1
	ACCUMULATED		
	ADDED	(DIFFERENT FORM APPEARS IN THE TEXT)	
	EXPANDED	(DIFFERENT FORM APPEARS IN THE TEXT)	
	INCREASED		
	LESSENED	(DIFFERENT FORM APPEARS IN THE TEXT)	
	REDUCED		
	--------REINFORCED	(DIFFERENT FORM APPEARS IN THE TEXT)	
131	ALSTRALIA		1
132	ALSTRIA'S		5
	ALSTRIA		
133	ALTHORITIES		4
	ALTHOR		
	ALTHORS		
134	ALTOMOBILE		1
135	AVAILAHILITY		2
	AVAILABLE		
136	AVERAGE		3
137	AVOIDED		3
	AVOID		
138	BACKWARD		2
139	BACTERICLOGICAL		1
140	BAGHDAD		1
141	BALANCE		2
142	BALDWIN		1
143	BALKAN		4
	BALKANS		
144	BALLISTIC		1
145	BARREN		1
146	BASED		35
	BASE		
	BASES		
147	BASICALLY		33
	BASIC		
	BASIS		
148	BATTLEFIELD		11
	BATTLEFIELDS		
	BATTLE		
	BATTLES		
149	BEARING		5
150	BEAR		2
	BEARS		
151	BECAME		9
	BECOME		
	BECOMES		
152	BEGAN		15
	BEGINNING		
	BEGIN		
	BEGUN		
153	BEHIND		3
154	BELGIUM		2

155	BELIEVE	
156	BELLIGERENT	8
	BELLIGERENTS,	
	BELLIGERENTS	
157	BELOW	
158	BENEFIT	1
	BENEFITS	6
159	BERNHARDI	
160	BESIDES	1
161	BETRAY	1
162	BEYOND	1
163	BIG	3
164	BILATERAL	1
165	BIOLOGICAL	1
166	BIRTH	2
167	BISMARCK	2
168	BITTER	1
169	BLACK	2
170	BLAZE	1
171	BLITZKRIEG	1
172	BLOC	1
	BLOCS	10
	BLOS	
173	BLOOD	
174	BLOW	2
	BLOWS	7
175	BLUE	
176	BODIES	1
	BODY	3
177	BOMBED	
	BOMBER	5
	BOMBS	
178	BOOK	
179	BORDER	2
	BORDERS	2
180	BORN	
181	BOSPORUS	1
182	BOUNDARIES	1
183	BOURBON	3
184	BOURGEOISIE	1
185	BOURGEOIS	3
186	BRANCHES	35
	BRANCH	16
187	BRAVERY	
188	BREAK	1
	BREAKS	3
189	BREAKTHROUGH	
190	BREWING	1
191	BRIEF-EXPOSITION	1
192	BRIEFLY	1
	BRIEF	3
193	BRING	
194	BRINKMANSHIP	5
195	BRITAIN	1
	BRITISH	17
196	BROUGHT	
197	BT	4
198	BUDGET	1
199	BUILDING	2
	BUILD	5

	BLILI	
200	BLILD-UP	1
201	BLLGARIA	2
202	BUSINESSMEN	
	BUSINESS	1
203	CAESAR	1
204	CALAMITY	4
205	CALCULATING	
	CALCULATIONS	3
206	CALLED	
	CALL	
	CALLS	1
207	CAME	2
208	CAMPAIGN	
	CAMPAIGNS	5
209	CAMP	1
210	CANADA	16
211	CANNOT	7
212	CAPABILITIES	
	CAPABILITY	
	CAPABLE	5
213	CAPACITIES	
	CAPACITY	34
214	CAPITALISM	
	CAPITALISTIC	
	CAPITALIST	
	CAPITALISTS	
	CAPITAL	3
215	CAPTURE	2
216	CAREFUL	1
217	CARPATHIAN	7
218	CARRIED	
	CARRYING	
	CARRY	11
219	CASE	
	CASES	1
220	CAST	3
221	CATASTROPHE	
	CATASTROPHIC	2
222	CATEGORIES	5
223	CAUSED	
	CAUSE	1
224	CAVALRY	1
225	CEASE	3
226	CENTERS	
	CENTRALIZATION	
	CENTRAL	2
227	CENTO	9
228	CENTURY	10
229	CERTAIN	1
230	CHANCELLOR	4
231	CHANCE	
	CHANCES	26
232	CHANGED	
	CHANGE	
	CHANGES	
	CHANGING	
233	*CHANNELED	1
	* --------ALLOCATED (DIFFERENT FORM APPEARS IN THE TEXT)	1
	DISPOSE (DIFFERENT FORM APPEARS IN THE TEXT)	

		DIRECTED GROUP METER	(DIFFERENT FORM APPEARS IN THE TEXT)	
234	CHAOTIC			1
235	CHAPTER			1
236	CHARACTERISTIC			12
	CHARACTERISTICS			
	CHARACTERIZED			
	CHARACTER			
237	CHEMICAL			
238	CHEMISTRY			1
239	CHESTNUTS			1
240	CHIANG			1
241	CHICAGO			1
242	CHOICE			2
243	CHRONICLERS			1
244	CHURCHES			1
245	CIRCLES			1
246	CIRCULATE			4
247	CIRCUMSTANCES			1
248	CITED			2
	CITE			2
249	CIVIL			
250	CLAIM			8
251	CLARIFICATION			1
	CLARIFIED			2
252	CLASSES			
	CLASSICAL			23
	CLASSLESS			
	CLASS			
253	CLASSIFICATION			
	CLASSIFYING			3
254	CLAUSEWITZ'S			
	CLAUSEWITZ			2
255	CLEARLY			
	CLEAR			11
256	CLOSED			
	CLOSELY			11
	CLOSER			
	CLOSE			
257	CCALITION			
	CCALITIONS			20
258	CCAL			
259	CCASTAL			4
260	CCERCION			2
261	CCEXISTENCE			3
262	CC-OPERATION			1
263	CC-ORDINATE			2
	CC-ORDINATING			6
	CC-ORDINATION			
264	*CCLD			
	--------HOT			1
265	CCLONIALISM			
	CCLONIAL			2
266	*CCLORED			
		AFFECTED		1
267	CCLOSSAL			
268	CCMBATING			2
	CCMBAT			27
269	CCMBINATION			
				3

	COMBINED			
	COMBINING			
270	COME			3
	COMES			
271	COMMAND			8
	COMMANDS			
272	COMMENSURATE			1
273	COMMIT			1
274	COMMON			2
275	COMMUNICATION			10
	COMMUNICATIONS			
276	COMMUNISM			10
	COMMUNIST			
277	COMPANIES			7
	COMPANY			
278	COMPARATIVE			1
279	COMPARED			4
	COMPARISON			
280	*COMPELLED			2
		COERCED	(DIFFERENT FORM APPEARS IN THE TEXT)	116
		FORCED		
		--------AUTHORITY	(DIFFERENT FORM APPEARS IN THE TEXT)	
		CONTROL		
		--------HELPLESS	(DIFFERENT FORM APPEARS IN THE TEXT)	
		MIGHT		
		POWERFUL	(DIFFERENT FORM APPEARS IN THE TEXT)	
		POWERLESS	(DIFFERENT FORM APPEARS IN THE TEXT)	
		POWER		
		STRENGTH		
		WEAK		
	--------NECESSITATED			
		OBLIGED	(DIFFERENT FORM APPEARS IN THE TEXT)	
		PRESSED	(DIFFERENT FORM APPEARS IN THE TEXT)	
281	COMPETITION			4
282	COMPILED			1
283	COMPLETED			23
	COMPLETELY			
	COMPLETE			
284	COMPLEXES			4
	COMPLEXITY			
	COMPLEX			
285	COMPONENT			2
286	COMPOSED			8
	COMPOSITION			
287	*COMPREHENSIVE			3
		WIDE		
288	COMPRISE			4
	COMPRISES			
	COMPRISING			
289	COMPROMISES			2
290	*CONCEAL			2
291	CONCENTRATED			17
	CONCENTRATE			
	CONCENTRATING			
	CONCENTRATION			
	CONCENTRATIONS			
292	CONCEPTION			36
	CONCEPT			
	CONCEPTS			
	CONCEPTUAL			

293	CCNCERNED			3
	CCNCERN			
	CCNCERNS			
294	CCNCESSIONS			
295	CCNCLUDED			2
	CCNCLUDING			10
	CCNCLUSION			
	CCNCLUSIONS			
296	CCNCRETE			
297	CCNCURRENCE			6
	CCNCUR			2
298	CCNDITION			
	CCNDITICNS			35
299	CCNDUCTED			45
	CCNDUCTING			
	CCNDUCT			
300	CCNFERENCES			
301	*CCNFINED			1
		LIMITED		1
302	CCNFIRMED			
	CCNFIRM			7
	CCNFIRMS			
303	CCNFORMITY			
	CCNFORM			2
304	CCNFRONTED			
305	CCNFUCIUS			1
306	CCNGRESSES			1
	CCNGRESS			3
307	CCNJUNCTION			
308	CCNNECTION			2
309	CCNQUER			2
	CCNQUEST			2
310	CCNSIDERABLE			
	CCNSIDERABLY			6
311	CCNSIDERATION			
	CCNSIDERATIONS			28
	CCNSIDERED			
	CCNSIDERING			
	CCNSIDER			
	CCNSIDERS			
312	CCNSISTED			
	CCNSISTENCY			10
	CCNSISTENT			
	CCNSISTING			
	CCNSIST			
	CCNSISTS			
313	*CCNSOLIDATE			
		HOLD		1
		UNITE		
			ACCORD	76
			AGREEMENT	
			DISAGREE	(DIFFERENT FORM APPEARS IN THE TEXT)
			SOLIDARITY	
314	CCNSTANTINOPLE			
315	CCNSTANTLY			1
	CCNSTANT			10
316	CCNSTRUCTION			
317	CCNSUMERS			6
318	CCNTAINMENT			1
319	CCNTENT			1
				8

320	CONTEXT			1
321	CONTINENTAL			6
	CONTINENT			
	CONTINENTS			
322	CONTINUAL			1
323	CONTINUATION			2
324	CONTINUOUS			2
325	CONTRADICTIONS			9
326	CONTRARY			2
327	*CONTRAST			1
		COMPARISON		
		DIFFERENCE		
328	CONTRIBUTED			3
	CONTRIBUTE			
	CONTRIBUTIONS			
329	CONTROL			3
	CONTROLS			
330	CONVENTIONAL			1
331	CONVERSION			1
332	CONVINCED			2
	CONVINCE			
333	CORRECTLY			3
	CORRECT			
334	CORRELATION			1
335	CORRESPONDENT			4
	CORRESPOND			
336	CORRESPONDINGLY			2
337	COST			2
338	COUNCIL			1
339	COUNTERATTACK			1
340	COUNTRIES'			99
	COUNTRIES			
	COUNTRY'S			
	COUNTRY			
		COLONY	(DIFFERENT FORM APPEARS IN THE TEXT)	
		CONTROL		
		DICTATION		
		DOMINATION		
		INFLUENCE		
		MASTERY		
		PATRIOTIC		
		POLITY	(DIFFERENT FORM APPEARS IN THE TEXT)	152
		DESIGN		
		OUTLINE		
		PLAN		
		PROGRAM		
		POWER		
	---------	PRINCIPALITY	(DIFFERENT FORM APPEARS IN THE TEXT)	
		REALM		
		SOVEREIGNTY		
		STATE		139
		CONDITION		
		COUNTRY		
		GOVERNMENT		
		REALM		
341	COUNT			1
342	COURAGE			1
343	COURSE			5
344	CPSU			1
345	CREATED			24

	CREATE		
	CREATES		
	CREATING		
	CREATION		
346	CRITERION		
347	CRITICAL		1
348	CRUSHING		1
349	CULT		2
	CULTURAL		7
	CULTURE		
350	CURRENT		
351	CUSTOMS		2
352	CYBERNETICS		1
353	DARDANELLES		1
354	DATA		1
355	DAY		1
	DAYS		5
356	*DEADLY		
	BLOOD		1
	DETRIMENTAL		
	--------HARMFUL	(DIFFERENT FORM APPEARS IN THE TEXT)	
	KILLING		
	--------UNHEALTHY		
357	DEALING		
	DEAL		7
	DEALS		
	DEALT		
358	DEAN		
359	DECEMBER		1
360	DECISIONS		2
	DECISIVELY		18
	DECISIVENESS		
	DECISIVE		
361	DECLARATION		
	DECLARATIONS		12
	DECLARED		
	DECLARES		
	DECLARING		
362	*DECLINE		
	ADD		1
*	--------AUGMENT	(DIFFERENT FORM APPEARS IN THE TEXT)	
	ACCUMULATED	(DIFFERENT FORM APPEARS IN THE TEXT)	1
	ADDED		
	EXPANDED	(DIFFERENT FORM APPEARS IN THE TEXT)	
	INCREASED	(DIFFERENT FORM APPEARS IN THE TEXT)	
	LESSENED		
	REDUCED	(DIFFERENT FORM APPEARS IN THE TEXT)	
	--------REINFORCED	(DIFFERENT FORM APPEARS IN THE TEXT)	
	DEVELOP		73
	ADVANCE		
	--------AGE		
*	--------AUGMENT	(DIFFERENT FORM APPEARS IN THE TEXT)	1
	ACCUMULATED		
	ADDED	(DIFFERENT FORM APPEARS IN THE TEXT)	
	EXPANDED	(DIFFERENT FORM APPEARS IN THE TEXT)	
	INCREASED		
	LESSENED	(DIFFERENT FORM APPEARS IN THE TEXT)	
	REDUCED		
	--------REINFORCED	(DIFFERENT FORM APPEARS IN THE TEXT)	
	--------DECLINE		

```
                EVOLUTION                                   (DIFFERENT FORM APPEARS IN THE TEXT)
                EVOLVE                                      (DIFFERENT FORM APPEARS IN THE TEXT)
                PROGRESS
        --------RIPEN                                       (DIFFERENT FORM APPEARS IN THE TEXT)
        GROW                                                (DIFFERENT FORM APPEARS IN THE TEXT)
        INCREASE
        LESSEN                                              (DIFFERENT FORM APPEARS IN THE TEXT)
*       DECREASE                                            (DIFFERENT FORM APPEARS IN THE TEXT)      0
                (PRECEDING PRIMARY WORD DOES NOT APPEAR IN THIS SECTION)
                ADD                                         (DIFFERENT FORM APPEARS IN THE TEXT)
*       --------AUGMENT                                    (DIFFERENT FORM APPEARS IN THE TEXT)      1
                        ACCUMULATED
                        ADDED                               (DIFFERENT FORM APPEARS IN THE TEXT)
                        EXPANDED                            (DIFFERENT FORM APPEARS IN THE TEXT)
                        INCREASED
                        LESSENED                            (DIFFERENT FORM APPEARS IN THE TEXT)
                        REDUCED
                --------REINFORCED                          (DIFFERENT FORM APPEARS IN THE TEXT)
        --------DECLINE                                                                             73
                DEVELOP
                        ADVANCE
                --------AGE
*               --------AUGMENT                            (DIFFERENT FORM APPEARS IN THE TEXT)      1
                                ACCUMULATED
                                ADDED                       (DIFFERENT FORM APPEARS IN THE TEXT)
                                EXPANDED                    (DIFFERENT FORM APPEARS IN THE TEXT)
                                INCREASED
                                LESSENED                    (DIFFERENT FORM APPEARS IN THE TEXT)
                                REDUCED
                        --------REINFORCED                  (DIFFERENT FORM APPEARS IN THE TEXT)
                --------DECLINE
                        EVOLUTION                           (DIFFERENT FORM APPEARS IN THE TEXT)
                        EVOLVE                              (DIFFERENT FORM APPEARS IN THE TEXT)
                        PROGRESS
                --------RIPEN                               (DIFFERENT FORM APPEARS IN THE TEXT)
                GROW                                        (DIFFERENT FORM APPEARS IN THE TEXT)
                INCREASE
                LESSEN                                      (DIFFERENT FORM APPEARS IN THE TEXT)
363 *DECREE                                                                                           1
                ACT
                COMMAND
                FORMULATE
                LAW
                ORDER
                RULE
364   DEEMS                                                                                           1
365   DEEP                                                                                            1
366   DEFEAT                                                                                         11
      DEFEATS
367   DEFENDING                                                                                       1
368   DEFENSE-INDUSTRI                                                                                1
369   DEFENSE                                                                                        29
      DEFENSIVE
370   DEFINED                                                                                        19
      DEFINES
      DEFINING
      DEFINITE
      DEFINITION
      DEFINITIVE
371 *DEFLECT                                                                                          1
```

--------SHIFT
TURN
DIVERT (DIFFERENT FORM APPEARS IN THE TEXT) 0
(PRECEDING PRIMARY WORD DOES NOT APPEAR IN THIS SECTION)
--------DEFLECT
--------SHIFT
TURN

372 DEGREE
373 DELAYED 5
DELAYING 2
374 *DELIBERATE
INTEND 1
375 DELIVERED (DIFFERENT FORM APPEARS IN THE TEXT)
DELIVERING 8
DELIVER
DELIVERY
376 DELL'-ARTE-DELLA
377 *DELUSIONS 1
--------ILLUSION 1
378 DEMANDED (DIFFERENT FORM APPEARS IN THE TEXT)
DEMANDS 4
379 DEMOCRATIC
380 DEMONSTRATED 2
DEMONSTRATION 2
381 DEMORALIZATION
DEMORALIZING 4
382 DENMARK
383 DENSITIES 1
DENSITY 2
384 DENY
385 DEPENDED 3
DEPENDENCE 26
DEPENDENT
DEPEND
DEPENDS
386 DEPLOYMENT
DEPLOY 12
387 DEPOSITS
388 DEPTH 1
389 DERIVE 3
390 DESCRIBED 1
DESCRIPTION 3
391 DESIGN
392 DESIRE 1
DESIRES 2
393 DESPITE
394 DESTROYED 1
DESTROY 15
DESTRUCTION
DESTRUCTIVE
395 DETAILED
DETAIL 2
396 DETERMINATION
DETERMINED 35
DETERMINE
DETERMINES
DETERMINING
397 DETERRENCE
398 *DETONATE 1
399 DETRIMENT 1
2

```
400   DEVELOPED                                                                                    73
      DEVELOPING
      DEVELOPMENT
      DEVELOP
                ADVANCE
        --------AGE
    *   --------AUGMENT                           (DIFFERENT FORM APPEARS IN THE TEXT)           1
                        ACCUMULATED
                        ADDED                     (DIFFERENT FORM APPEARS IN THE TEXT)
                        EXPANDED                  (DIFFERENT FORM APPEARS IN THE TEXT)
                        INCREASED
                        LESSENED                  (DIFFERENT FORM APPEARS IN THE TEXT)
                        REDUCED
                --------REINFORCED                (DIFFERENT FORM APPEARS IN THE TEXT)
    *   --------DECLINE                                                                          1
                        ADD                       (DIFFERENT FORM APPEARS IN THE TEXT)
    *           --------AUGMENT                   (DIFFERENT FORM APPEARS IN THE TEXT)           1
                                ACCUMULATED
                                ADDED             (DIFFERENT FORM APPEARS IN THE TEXT)
                                EXPANDED          (DIFFERENT FORM APPEARS IN THE TEXT)
                                INCREASED
                                LESSENED          (DIFFERENT FORM APPEARS IN THE TEXT)
                                REDUCED
                        --------REINFORCED        (DIFFERENT FORM APPEARS IN THE TEXT)
                        DEVELOP
                        GROW                      (DIFFERENT FORM APPEARS IN THE TEXT)
                        INCREASE
                        LESSEN                    (DIFFERENT FORM APPEARS IN THE TEXT)
    *                   DECREASE                  (DIFFERENT FORM APPEARS IN THE TEXT)           0
                                (PRECEDING PRIMARY WORD DOES NOT APPEAR IN THIS SECTION)
                                ADD               (DIFFERENT FORM APPEARS IN THE TEXT)
    *                   --------AUGMENT           (DIFFERENT FORM APPEARS IN THE TEXT)           1
                                        ACCUMULATED
                                        ADDED     (DIFFERENT FORM APPEARS IN THE TEXT)
                                        EXPANDED  (DIFFERENT FORM APPEARS IN THE TEXT)
                                        INCREASED
                                        LESSENED  (DIFFERENT FORM APPEARS IN THE TEXT)
                                        REDUCED
                                --------REINFORCED (DIFFERENT FORM APPEARS IN THE TEXT)
                        --------DECLINE
                                DEVELOP
                                GROW              (DIFFERENT FORM APPEARS IN THE TEXT)
                                INCREASE
                                LESSEN            (DIFFERENT FORM APPEARS IN THE TEXT)
                EVOLUTION                         (DIFFERENT FORM APPEARS IN THE TEXT)
                EVOLVE                            (DIFFERENT FORM APPEARS IN THE TEXT)
                PROGRESS
        --------RIPEN                             (DIFFERENT FORM APPEARS IN THE TEXT)
      DEVELOPS
401   DEVIOUS                                                                                    1
402  *DEVISE                                                                                     2
                DESIGN
                INVENT                            (DIFFERENT FORM APPEARS IN THE TEXT)
                OUTLINE
                PLAN
403   DEVOTED                                                                                    4
      DEVOTE
      DEVOTES
404   DIALECTICAL                                                                                3
      DIALECTIC
```

405 DIALOGUE
406 DICTATE 1
DICTATION 4
407 DID
408 DIFFERENCES /
DIFFERENTIATION 16
DIFFERENTLY
DIFFERENT
DIFFER
DIFFERS
409 DIFFICULTIES
DIFFICULT 5
DIFFICULTY
410 DILLON-REED 1
411 DILLON 1
412 DIPLOMACY 16
DIPLOMATIC
DIPLOMATS
413 DIRECTED 30
DIRECTING
DIRECTION
DIRECTIONS
DIRECTLY
DIRECTOR
DIRECTORS
DIRECT
DIRECTS
414 DISAGREEMENTS 1
415 DISAPPEARED 1
416 DISARMAMENT 1
417 *DISARRAY 1

METHOD
ORDER
SYSTEM
* DISORDER (DIFFERENT FORM APPEARS IN THE TEXT) 0
(PRECEDING PRIMARY WORD DOES NOT APPEAR IN THIS SECTION)
* --------ARRANGE 2
DESIGN
* --------DEVISE 2
DESIGN
INVENT (DIFFERENT FORM APPEARS IN THE TEXT)
OUTLINE
PLAN
ORGANIZE
PLAN
--------SYSTEMATIC
--------DISARRAY
METHOD
ORDER
ORGANIZE
SYSTEM
* DISORGANIZED (DIFFERENT FORM APPEARS IN THE TEXT) 0
(PRECEDING PRIMARY WORD DOES NOT APPEAR IN THIS SECTION)
--------DISARRAY
METHOD
ORDER
--------ORGANIZED
SYSTEM
DISORDER (DIFFERENT FORM APPEARS IN THE TEXT)
* DISORGANIZED (DIFFERENT FORM APPEARS IN THE TEXT) 0

(PRECEDING PRIMARY WORD DOES NOT APPEAR IN THIS SECTION)

--------DISARRAY
METHOD
ORDER
--------ORGANIZED
SYSTEM
DISORDER (DIFFERENT FORM APPEARS IN THE TEXT) 0

(PRECEDING PRIMARY WORD DOES NOT APPEAR IN THIS SECTION)

* --------ARRANGE 2
DESIGN
* --------DEVISE 2
DESIGN
INVENT (DIFFERENT FORM APPEARS IN THE TEXT)
OUTLINE
PLAN
ORGANIZE
PLAN
--------SYSTEMATIC
--------DISARRAY
METHOD
ORDER
ORGANIZE
SYSTEM
DISORGANIZED (DIFFERENT FORM APPEARS IN THE TEXT)

418 *DISASTER 2
--------CALAMITY
CATASTROPHE
--------DISASTROUS
419 DISASTROUS 1
420 DISCIPLINE 6
DISCIPLINES
421 DISCOVERIES 5
DISCOVERY
422 DISCUSSED 7
DISCUSSES
DISCUSSING
DISCUSSIONS
DISCUSS
423 DISPOSAL 5
424 DISPUTES 2
425 DISRUPTED 3
DISRUPTION
426 DISTANCE 2
DISTANCES
427 DISTINCTIONS 1
428 DISTRIBUTE 6
DISTRIBUTION
429 *DISTRICTS 1
AREAS
LOCALITY (DIFFERENT FORM APPEARS IN THE TEXT)
REGION
SECTIONS
SECTOR
TERRITORY
ZONE
430 DIVERSITY 1
431 DIVIDED 4
DIVISION
DIVISIONS
432 DOCTRINE 10

433	DCCTRINES			
	DCES			7
434	DCING			1
435	DCMESTIC			2
436	DCMINANT			4
	DCMINATE			
437	DCMINATED			3
	DCMINATION			
438	DCNE			3
439	DC			10
440	*DCUBLE			1
		FALSE		
		---------TRUSTLESS	(DIFFERENT FORM APPEARS IN THE TEXT)	
441	DCUGLAS			1
442	DCWNFALL			3
	DCWN			
443	DRAWN			1
444	*DRILL			1
		DISCIPLINE		
		INSTRUCTION	(DIFFERENT FORM APPEARS IN THE TEXT)	
		PRACTICE		
		PREPARATION		
		SCHOOLING	(DIFFERENT FORM APPEARS IN THE TEXT)	
		TRAINING		
445	DRIVE			2
	DRIVES			
446	DR.			2
447	DLE			2
448	DULLES			1
449	DURATION			2
450	DYNASTY			1

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)*

System Development Corporation
Santa Monica, California

2a. REPORT SECURITY CLASSIFICATION

Unclassified

2b. GROUP

3. REPORT TITLE

Stylistic Analysis Report On The Second Year Of Research

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

5. AUTHOR(S) *(Last name, first name, initial)*

Sedelow, Sally Y.

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
1 March 1966	137	

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
Nonr-4427(00)	TM-1908/200/00
b. PROJECT NO	
c.	9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*
d.	

10. AVAILABILITY/LIMITATION NOTICES

This document has been cleared for open publication and may be disseminated by the Clearing House for Federal Scientific & Technical Information.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY

13. ABSTRACT

This document describes the current capabilities of the computer-based programs being developed for the analysis of style. The power of the programs for content analysis, semantic analysis and stylistic discrimination is evaluated on the basis of their analytical runs on natural language text. The value of such output for political and diplomatic negotitations, as well as for more long-range projects involving information retrieval and automatic language production, etc., is illustrated.

DD FORM 1 JAN 64 1473

Unclassified

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Stylistic Analysis Philco 2000 Language Natural Sort Text Index Computer-Based Programs						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2a. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7a. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8a. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, 8c, & 8d. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through _____."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through _____."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through _____."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS)*, *(S)*, *(C)*, or *(U)*

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, rules, and weights is optional.

System Development Corporation/2500 Colorado Avenue/Santa Monica, California 90406