

A U T O M A T E D L A N G U A G E A N A L Y S I S

1968 - 1969

Report on research for the period
March 1, 1968 - February 28, 1969

Sally Y. Sedelow, Principal Investigator
Departments of English and Computer & Information Science
University of North Carolina
at
Chapel Hill

The research reported herein was conducted under Contract N00014-
67-A-0321, Office of Naval Research, U.S. Navy, Task No. NR 348-005

Distribution of this Document is Unlimited

The research reported herein was conducted under Contract N0014-67-A-0321, Office of Naval Research, U.S. Navy, Task No. NR 348-005

at the
University of North Carolina
at Chapel Hill

A U T O M A T E D L A N G U A G E A N A L Y S I S

1968 - 1969

Report on research for the period
March 1, 1968 - February 28, 1969

Sally Y. Sedelow, Principal Investigator
Thomas Gerig, Consultant
Walter A. Sedelow, Jr., Consultant
Walter L. Smith, Consultant
H. William Buttlemann
William G. Hickok
Joan Peters
Larry Rosen
John B. Smith

The views, conclusions, or recommendations expressed in this document do not necessarily reflect the official views or policies of agencies of the United States Government.

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

Copyright © 1969, by Sally Yeates Sedelow

Reproduction of this document in whole or in part is
permitted for any purpose of the United
States Government.

1 March 1969

1

Automated Language Analysis

1968 - 1969

Report on Research for the Period
March 1, 1968 - February 28, 1969

Sally Yeates Sedelow, Principal Investigator

ABSTRACT

This report describes research directed toward unequivocally characterizing the language usage of any writer or speaker. During the past year, a PREFIX program has increased the root-grouping capability of the VIA package, which is a set of programs designed for content, or thematic, analysis of a written or spoken language unit. A ring-structure version of VIA which provides great search flexibility as well as the potential for extended explorations of semantic relationships, as revealed by interconnecting rings, has been further developed. Research on the nature of thesauri has continued and Roget's International Thesaurus has been keypunched to facilitate computer-aided research on its structure. A set of programs designed to show co-occurrence patterns has been implemented, as have procedures for producing non-verbal representations of language-usage patterns.

1 March 1969

3

PREFACE

This project has received support from many sources during the past year. The members of the Information Systems Staff of the Office of Naval Research have been unfailingly helpful whenever advice or administrative support was sought. The University of North Carolina at Chapel Hill has also provided administrative assistance. In addition, the Department of Computer and Information Science has provided office space and equipment as well as secretarial support for the research group and the Department of English has granted the principal investigator a reduced teaching load, freeing time which is partially allocated to this project. We also wish to extend our thanks to all the individuals who have contributed, in one way or another, to the furtherance of this research effort.

TABLE OF CONTENTS

	Page
Preface	3
I. Survey of the Computer-Aided Language Analysis.....	
Project March 1, 1968 - February 28, 1969	7
A. Introduction.....	7
B. Work During the Past Year.....	10
1. PREFIX.....	12
2. Principal Component Analysis: A Discussion.....	26
3. Thesaurus Research.....	35
4. Ring-Structure Version of VIA.....	40
C. Plans for Further Research and Development.....	49
II. Program Documentation.....	54
A. Ring-Structure VIA.....	54
B. PREFIX.....	82
C. CONTEXT.....	92
III. Professional Activities of Project Personnel.....	116
IV. Appendices.....	123
A. VIA System Information Flow Through the SUFFIX Program.....	123
B. Ring-Structure VIA Program Listings.....	125
C. Ring-Structure VIA Utility Programs.....	183
D. List-Structure VIA Thesaurus Program Listing and Documentation.....	199
E. MAPTEXT Listing and Sample Output.....	231
F. PREFIX Program and Table Listing.....	245
G. CONTEXT Listings.....	275