

ATOMATED ANALYSIS OF LANGUAGE  
STYLE AND STRUCTURE

1969 - 1970

Report on research for the period  
March 1, 1969 - August 31, 1970

The Research reported herein was conducted under Contract  
N00014-67-A-0321-001, Office of Naval Research, U. S. Navy,  
Task No. NR348-005

at the  
University of North Carolina  
at Chapel Hill

Distribution of this Document is Unlimited

The research reported herein was conducted under Contract  
N00014-67-A-0321-001, Office of Naval Research, U. S. Navy,  
Task No. NR348-005

at the

University of North Carolina  
at Chapel Hill

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

1969-1970

Report on research for the period  
March 1, 1969 - August 31, 1970

Sally Y. Sedelow, Principal Investigator  
Martin Dillon, Consultant  
Gerald Fisher, Consultant  
Walter Sedelow, Consultant  
Walter Smith, Consultant  
H. William Buttelmann  
John B. Smith  
David Wagner

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 © 1970, by Sally Yeates Sedelow

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

AUTOMATED ANALYSIS OF LANGUAGE

STYLE AND STRUCTURE

1969 - 1970

Report on Research for the Period  
March 1, 1969 - August 31, 1970

Sally Yeates Sedelow, Principal Investigator

ABSTRACT

This report summarizes work directed toward comprehensive computer-based characterization of natural language text. A general-purpose thesaurus has been keypunched and proofread, and corrections are now being made; this work is preliminary to an effort to determine the structure of the thesaurus and, in general, to construct parametric models of thesauri. Preliminary approaches to the design of a statistical package for natural language analysis are also described; these approaches include a program to provide data formats for any type of language analysis. A User's Manual for the PL/I, S/360 list-structure and ring-structure VIA packages is also included in this report.

## PREFACE

I would like to thank the staff of the Information Systems Branch, Office of Naval Research, for their unfailing helpfulness on matters connected with this contract. Thanks are also due to the Departments of English and Computer and Information Science for a reduced teaching load, permitting work on this research, and to the Office of Research Administration for assistance with administrative detail. I am also grateful to the secretaries in the Department of Computer and Information Science for getting the report into reproducible form and overseeing its circulation.

## TABLE OF CONTENTS

	Page
Preface . . . . .	3
I. Survey of the Automated Analysis of Language Project, March 1, 1969 - August 31, 1970 . . . . .	7
A. Introduction . . . . .	7
B. Summary of First Year's Work . . . . .	8
C. Plans for Future . . . . .	13
D. Extended Discussions . . . . .	15
1. Thesaurus Research . . . . .	15
2. Preliminary Design Considerations of a Language Analysis Support Package . . . . .	18
3. Design Requirements for a Generalized Data- Formatting Package . . . . .	38
4. A Design for a General Statistical Analyzer for Natural Language Texts . . . . .	75
II. Program Documentation . . . . .	85
A. Updated User's Manual for PL/I, S/360 VIA . . . . .	85
1. Part I . . . . .	89
2. Part II . . . . .	100
B. PREFIX--Revised Version . . . . .	131
III. Professional Activities of Project Personnel . . . . .	136
IV. Appendix: Brief Description and Listing for Preliminary Data-Formatting Program . . . . .	148