Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

I 	М	P	L	E	M	Ε	N	T	A	T	I	0	N
	v	I	A			E	N	G	L	I	S	H	

Owen D. Kessell

A thesis presented in partial fulfilment of the requirement for the degree of

Master of Science in Computer Science

at

Massey University

December 1976.

ABSTRACT

In this thesis, synthesis of COBOL programs is discussed. The programs are generated with the aid of an interactive Natural Language dialogue.

Reasons for and against the use of English as a general programming aid are discussed, also the use of English in program synthesis is discussed.

The major portion of this thesis describes the design of the system known as CLIVE. The discussion illustrates the relative ease in which COBOL programs can be generated by using ordinary English responses.

ACKNOWLEDGEMENTS

In presenting this thesis I would like to take this opportunity to express my thanks to the following people:

First and foremost to my supervisors, Brian Carpenter and Lindsay Groves, whose continual guidance and criticizm has been most helpful.

To the various members of the Computer Science Department and Computer Unit at Massey University who, through assorted discussions, assisted in the gathering of information contained in this thesis; especially to Neil James, for his assistance with details regarding the implementation of CLIVE.

Finally to my wife Margaret, for her continual understanding and patience.

Massey University December 1976 Owen D. Kessell

TABLE OF CONTENTS

		Page			
Chapter 1	Introduction	1			
Chanton 2	The Dessens For Such & Sustam	7			
Chapter 2	The Reasons for Such A System	2			
2.1	Marginal and Economic Reasons	3			
2.2	Natural Reasons	5			
2.3	Error Reduction	5			
2.4	Conclusion	6			
Chapter 3	Overview	7			
3.1	Program Synthesis	7			
3.2	COBOL Programs	8			
3.3	Question and Answer Dialogue	9			
3.4	A Sample Dialogue	15			
3.5	Structure of the System	19			
3.6	Readability	20			
3.7	Other COBOL Source Code Generators	22			
	3.7.1 SURGE	22			
	3.7.2 Computer Aided Program Production	23			
3.8	CLIVE Re-visited	24			
Chapter 4	The Input Module	26			
4.1	Introduction	26			
4.2	Ambiguity	27			
4.3	Natural Language Understanding Systems				
4.4	The Specific Approach	32			
	4.4.1 LUNAR	32			
	4.4.2 GENIE	33			

4.5	The Problem To Be Solved		34
	4.5.1 Question Boundaries		37
	4.5.2 Idioms		39
	4.5.3 Common Words and Punctuation		40
× *	4.5.4 Synonyms		41
	4.5.5 Suffixes and Prefixes		43
	4.5.6 Analysis Finished		44
	21		
Chapter 5	The Data-File Module	38	45
5.1	Introduction		45
5.2	The Identification Division		46
5.3	The Environment Division		46
5.4	The Data Division		47
	5.4.1 The Input File		48
	5.4.2 The Data To Be Used		50
	5.4.3 The Output File		51
5.5	The Working-Storage Section		53
	5.5.1 Titles		54
	5.5.2 Headings		55
	5.5.3 Tables		55
	5.5.4 Counts and Total Lines		56
5.6	Pictures		58
Chapter 6	The Procedure Module		61
6.1	Introduction		61
6.2	Standard Logic		62
6.3	The Question		64
6.4	Calculations		67
6.5	Program Completed	6.2	69
	<u>v</u>		

•

Chapter 7	Ancillary Modules	70
7.1	Introduction	70
7.2	Spelling Mistakes	70
7.3	Descriptive Names	72
7.4	Data Checking	73
7.5	The Assumption Module	74
Chapter 8	Conclusion	76
Appendix I	A Listing From A Session With The Current Implementation Of CLIVE	78
Appendix II	<u>A Listing Of The Results That Were Produced</u> From The Session Of Appendix I	80
Appendix III	I <u>A Listing Of The Program That Was Produced</u> From The Session Of Appendix I	81

References and Bibliography

86

Figure							Page
1	Structured	Outline	Of	CLIV	IΕ		19
2	Structured	Diagram	Of	The	Input	Analysis	36

CHAPTER 1.

INTRODUCTION.

"Where shall I begin, please your Majesty?" he asked. "Begin at the beginning", the King said gravely, "and go on till you come to the end: then stop."

> Lewis Carroll. "Through the Looking Glass".

CLIVE, which is an acronym for <u>COBOL</u> Language <u>Implementation</u> <u>Via English</u>, is a system that generates COBOL source programs from a Natural Language discourse.

CLIVE is used in a interactive manner; the users responses forming the basis upon which the desired program is generated.

Chapter Two discusses the justification for the use of Natural Language as an input medium. It is not the authors intention to advocate using English as a programming language but to advocate the use of it as a programming <u>aid</u>. Chapter Three discusses the approach taken towards program synthesis. A description of a sample dialogue is given to illustrate the types of responses that the user and CLIVE may give, and the chapter concludes with a description of CLIVE's overall construction.

1

Chapters Four, Five and Six discuss the major components of CLIVE: How they generate the appropriate section of COBOL code and how the user may respond to convey his information. Chapter Seven discusses briefly some modules that aren't strictly neccessary but are added to improve CLIVE's performance.

Chapter Eight is a brief conclusion about CLIVE and the overall philosophy behind it.

The appendices contain a sample session and results, using the current implementation of CLIVE which was done on the B 6700 at Massey University.