

AIN SHAMS UNIVERSITY
FACULTY OF ENGINEERING

NATURAL LANGUAGE PROCESSING
APPLIED TO ARABIC LANGUAGE

By

ENG. SALWA ELSAYED HAMADA

A Thesis

**Submitted in partial fulfillment of
the requirements of Master Degree
in Computer and Control Engineering**

Supervised by

Prof. Dr. Gamal M. Aly
Faculty of Engineering
Ain Shams University

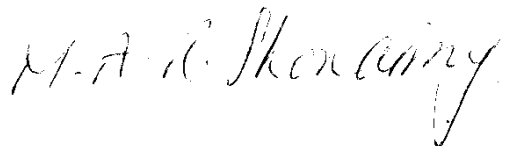
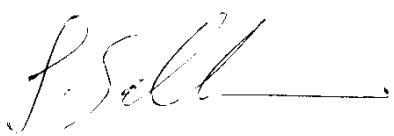
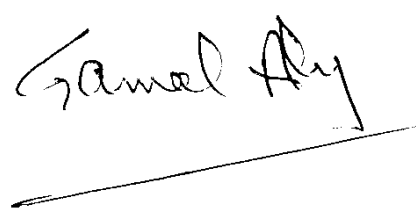
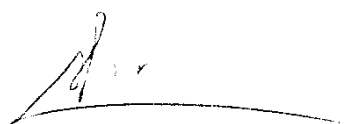
Prof. Dr. Nadia H. Hegazi
Electronic Research Institute
National Research Center

Cairo - 1989

بِسْمِ
اللّٰهِ الرَّحْمٰنِ الرَّحِیْمِ



Examiners Committee

<u>Name, Title & Affiliation</u>	<u>Signature</u>
1. Prof.Dr. M.A.R.Ghonaimy Faculty of Engineering Ain Shams University	
2. Prof. Dr. F.H. Saleh Faculty of Engineering Cairo University	
3. Prof. Dr. G.M. Aly Faculty of Engineering Ain Shams University	
4. Prof. Dr. N.H.Hegazi Electronic Research Institute National Research Center	

STATEMENT

This dissertation is submitted to Ain Shams University
for the degree of Master in Computer and control Engineering.

The work included in this thesis was carried out by
the author in the Department of Computer and Control, Ain Shams
University, from 6/11 1984 to 13/11/ 1989

No part of this thesis has been submitted for a degree
or a qualification at any other University or Institution.

Date : 13 / 11 /1989

Signature: *Salwa Hamada*

Name : Salwa Elsiad Hamada

TO MY ARABIC NATION

Acknowledgement

It is my pleasure to record my sincere thanks to professor Dr.Gamal M.Aly ,professor of automatic Control and Computers Ain Shams University ,for his supervision , constant courtesy and guidance throughout this research work.

It is my pleasure to present my deep gratitude to my supervisor Prof. Dr. Nadia Hamed Hegazi , Head of the consultation laboratory for Scientific Computation and statistics. (NRC) for her kind encouragement, helpful guidance and valuable advices through the course of this research . Her helpful assistance, discussion of results and correction of this manuscript is greatly acknowledged .

The author also greatly acknowledges Dr.Ehsan Mostafa Abed, researcher in Consultation Lab. for scientific computation and statistics, for her help and useful suggestions

I should also thank Prof. Dr. Kamal M. A. Bishr, Professor of Linguistics and Phonetics, Faculty of Dar Elaloum, Cairo University for his help in revising the arabic part of this thesis.

My beloved parents also deserve a word of appreciation for their encouragement and moral support.

Lastly but by no means least, I wish to record my deepest gratitude and thanks to my husband and children without their patience and continuous encouragement, this work would have never been possible.

ABSTRACT

Natural language processing is now a topic of interest in diverse academic and applied fields. Many useful tasks could be performed if the computer could be programmed to extract the content of a message that has been delivered to it or stored in it in natural language form.

With the advent of the 5th generation computer a great deal of work is expected in the area of computational linguistics. This is being done for Latin speaking languages, as for Arabic, this part is lacking.

The work in this thesis covers a missing part for Arabic, that is Natural Arabic Language Understanding intermixing Morphology, syntax and semantics.

Design of "ARAB" which is a natural understanding system was done.

The work included the design of a Morphological analyzer, a syntactic analyzer and intermixing it with a semantic interface using PROLOG as a programming language.

This system could be considered as a nucleus to an Expert system, Educational system, Question - Answering system, Translating system, and Data - Retrieval system.

TABLE OF CONTENTS

CHAPTER ONE : INTRODUCTION	Page
1.1 MOTIVATION	1
1.2 OBJECTIVE OF THE THESIS.....	2
1.3 PROPERTIES OF THE PROPOSED SYSTEM ..	3
1.4 STATEMENT OF THE PROBLEM ;SOLUTION APPROACHES	4
1.5 SCOPE OF WORK	6
CHAPTER TWO : NATURAL LANGUAGE PROCESSING	
2.1 INTRODUCTION	8
2.2 PROPERTIES OF LANGUAGES	10
2.3 GENERAL ORGANIZATION OF A NATURAL LANGUAGE PROCESSOR	12
2.4 MORPHOLOGICAL PHASE (LEXICAL ANALYSIS) .	12
2.5 SYNTAX ANALYSIS	15
2.6 EXPRESSING THE RULES OF SYNTAX	16
2.7 DIAGRAMMING SENTENCES	16
2.8 WHY DO WE CARE ABOUT SENTENCE STRUCTURE.	19
2.9 LOGIC PROGRAMMING	19
2.10 NON - PROCEDURAL PROGRAMMING	20
2.11 BACK - TRACKING	22
2.12 ARABIC (NL) UNDERSTANDING MODEL ...	23
2.13 AN EXPERT SYSTEM FOR UNDERSTANDING ARABIC SENTENCE	24
	Page

2.14 A NATURAL ARABIC LANGUAGE	25
UNDERSTANDING SYSTEM (NALUS)	
2.15 COMPARISON	26
2.15.1 EXAMPLES FOR POSITION أسم الفاعل	
أسم المفعول " IN A SENTENCE	33
2.15.2 EXAMPLE OF AMBIGUOUS SENTENCES	34
2.15.3 EXAMPLES OF AMBIGUOUS WORDS ...	36
CHAPTER THREE : PARSING	
3.1 INTRODUCTION	37
3.2 GENERAL ISSUES IN PARSER DESIGN	38
3.3 AUGMENTED TRANSITION NETWORKS (ATN)..	41
3.4 CONTEXT - FREE GRAMMARS	43
3.5 CONTEXT - FREE LANGUAGE	45
3.6 SYNTACTIC PARSING	46
3.7 REQUIREMENTS OF A NATURAL LANGUAGE PARSER	46
CHAPTER FOUR : DEVELOPMENT OF SYNTAX ANALYZER FOR NATURAL ARABIC TEXT	
4.1 INTRODUCTION	49
4.2 METHODOLOGY	51
4.3 SYNTACTIC PARSER INTERLEAVED WITH A SEMANTIC INFERENCE	56

CHAPTER FIVE : ARAB SYNTAX PARSER

5.1 INTRODUCTION	58
5.2 USER INTERFACE	58
5.3 MORPHOLOGICAL DICTIONARY	59
5.4 LEXICAL ANALYZER	65
5.5 SYNTACTIC PARSER	65
5.6 SEMANTIC INFERENCE.....	77
5.7 AUGMENTED TRANSITION NETWORKS (ATN).	81
5.8 BACUS - NAUR FORM (BNF) FOR ARABIC LANGUAGE	84
5.9 KNOWLEDGE BASE (KB) REPRESENTATION.	84
5.10 HOW TO UNDERSTAND AN ARABIC SENTENCE .	85
5.11 FINAL OUTPUT	86

CHAPTER SIX : CONCLUSIONS AND FUTURE WORKS

6.1 EVALUATION OF THE PROPOSED SYSTEM ...	88
6.2 CONCLUSIONS AND FUTURE WORK ...	91

LIST OF FIGURES

	Page
2.1 TREE - NOTATION FOR A DIAGRAMMED SENTENCE....	18
2.2 A PARSE TREE.....	18
3.1 PARSE TREE FOR " THE BOY IS CLEVER ".....	39
3.2 THE FINITE STATE TRANSITION DIAGRAM.....	42
3.3 A CONTEXT - FREE GRAMMAR.....	44
3.3 B EXPLANATION FOR THE ABBREVIATION.....	44
3.4 A CONTEXT - FREE DERIVATION FOR " على يحب الرياضه ".....	45
4.1 CLASSIFICATION OF ARABIC SENTENCE.....	53
4.2 OVERVIEW OF THE PROPOSED ARAB SYSTEM.....	57
5.1 RELATION BETWEEN MORPHOLOGICAL DICTIONARY AND LEXICAL ANALYZER.....	63 - 64
5.2 a PARSING OF THE SENTENCE " محمد المجتهد ناجح ومتفوق في المدرسه "....	69
5.2 b RELATION BETWEEN MORPHOLOGICAL , SYNTACTIC AND SEMANTIC ANALYSIS.....	70
5.3 EXAMPLE FOR PARSING A NOMINAL SENTENCE.....	72
5.4 EXAMPLE FOR PARSING A VERBAL SENTENCE.....	73
5.5 EXAMPLE FOR PARSING A TOOL SENTENCE.....	74
5.6 EXAMPLE OF A BI - COMPOSITE ARABIC SENTENCE	75

5.7	EXAMPLE OF A TRI - COMPOSITE ARABIC SENTENCE	76
5.8	STATE DIAGRAM FOR AN ARABIC NOMINAL SENTENCE	79
5.9	PHRASE MARKER FOR THE STATE DIAGRAM IN FIG. (5.8)	80
5.10	THE FINAL OUTPUT	87

APPENDICES

- APPENDIX (I) : B N F FOR ARABIC LANGUAGE.
- APPENDIX (II) : MONO - NOMINAL ARABIC SENTENCE.
- APPENDIX (III) : MONO - VERBAL ARABIC SENTENCE.
- APPENDIX (IV) : BI & TRI COMPOSITION ARABIC SENTENCE.
- APPENDIX (V) : EXAMPLES FOR PARSING ARABIC SENTENCES.
- APPENDIX (VI) : SYNTAX EXAMPLES FOR ARABIC SENTENCES.

LIST OF IMPORTANT SYMBOLES

ATN	AGUMENTED TRANSTION NETWORKS
BNF	BACUS-NAUR FORM
KB	KNOWLEDGE BASE
NALUS	NATURAL ARABIC UNDERSTANDING SYSTEM
NL	NATURAL LANGUGE
NLP	NATURAL LANGUAGE PROCESSING