

AIN SHAMS UNIVERSITY
FACULTY OF ENGINEERING

A MODEL FOR ARABIC SYNTAX AND MORPHOLOGY

By

SAMEH AHMED MOHAMED YOUSSEF

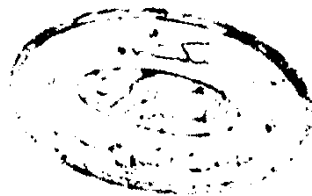
A Thesis

**Submitted in partial fulfillment for the
requirements of the Degree of M.Sc. in
Electrical Engineering (Computer)**

Supervised by

Prof. Dr. HUSSEIN I. SHAHEIN

Cairo - 1990



Examiners Committee

<u>Name, Title & Affiliation</u>	<u>Signature</u>
1 - Prof. Dr. F. H. Saleh Professor Faculty of Engineering Cairo University	
2 - Prof. Dr. M. A. R. Ghonaimy Professor and Chairman of Electronics and Computer Engineering Department Faculty of Engineering Ain Shams University	M. A. R. Ghonaimy
3 - Prof. Dr. H. I. Shahein Professor Faculty of Engineering Ain Shams University	H. I. Shahein

Date : / / 1990



STATEMENT

This dissertaion is submitted to Ain Shams University for the degree of Master of Science in Electrical Engineering.

The work included in this thesis was carried out by the author in the Department of Electronics and Computer, Ain Shams University, from October 1986 to October 1989.

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

Date : 27/1/1990

Signature : Samir Youssif

Name : Samir Ahmed Youssif

TO:

MY FATHER

and

MY MOTHER

Ain Shams University
Faculty of Engineering

Dept of : Electronics and Computer Engineering

Abstract of the M.Sc. Thesis submitted by:

SAMEH AHMED MOHAMED YOUSSEF

Title of Thesis: "A MODEL FOR ARABIC SYNTAX AND MORPHOLOGY"

Supervisor: PROF. HUSSEIN I. SHAHEIN

Registration Date: 10/11/1986 Examination Date:

Abstract: This thesis presents a model for Arabic syntax and morphology . A novel representation for lexicon structure is developed. This lexicon is completely designed from syntactic viewpoint. A classification for Arabic word has been achieved and syntactic features for all types of the word has been extracted. Furthermore, an integrated powerful Arabic morphological analyzer based on production system is developed. This morphological analyzer treats most of the disadvantages of previous models suggested before. In this model, a new technique called EC-hashing has been proposed to accelerate the analysis process, a complete view for realization rules has been suggested, the model is designed as a production system for flexible modification, and the output from this model simplifies the job of the syntax analyzer.

For Arabic syntax , A comprehensive model has been developed. This model has been implemented using augmented transition network grammars and the main ideas of systemic grammar. Also, algorithms have been proposed to handle deeply syntactic issues such as pronoun reference and conjunction. The models for syntax and morphology is implemented using prolog on personal computer.

ACKNOWLEDGMENT

I wish to express my deep gratitude and thanks to Professor Hussein I. Shahein for his generous help, constructive encouragement, kind advice, time and effort he devoted for supervision of this work as well as for his continuous discussion and creative directioning during all the stages of this work.

I would like also to thank my colleagues for their kind help.

I owe my family a lot for their priceless support, their encouragement and understanding during the hard work of this research.

CONTENTS

CHAPTER ONE INTRODUCTION

1.1	OVERVIEW	1
1.2	COMPUTER APPLICATIONS FOR NATURAL LANGUAGES.....	4
1.3	A CRITIQUE SURVEY ON PREVIOUS WORK	5
1.3.1	Arabic Morphology.....	5
1.3.2	Arabic Syntax.....	7
1.4	THESIS OBJECTIVE	9
1.5	THESIS SUMMARY	10

CHAPTER TWO SYNTACTIC MODELS

2.1	INTRODUCTION	11
2.2	CONTEXT-FREE GRAMMARS (CFGs)	12
2.2.1	The Inadequacies of Context-Free Grammars ..	14
2.3	DEFINITE CLAUSE GRAMMARS (DCGs)	14
2.3.1	The Use of Definite Clause Grammars	15
2.3.1.1	Building structures	15
2.3.1.2	Extra conditions	17
2.3.1.3	Context dependency	17
2.4	TRANSITION NETWORK GRAMMARS	19
2.4.1	Recursive Transition Networks	20
2.4.2	Augmented Transition Networks	22
2.5	COMPARATIVE ANALYSIS BETWEEN GRAMMARS	24
2.6	SYSTEMIC GRAMMAR	27
2.6.1	The Goal of Systemic Grammar	27
2.6.2	Classification	29

2.6.3	Syntax Resources	30
2.6.4	Choice Systems	31
2.6.4.1	System networks	32
2.6.5	Realization in Systemic Grammar	35

CHAPTER THREE ARABIC LEXICON

3.1	INTRODUCTION	37
3.2	THE CLASSIFICATION OF ARABIC WORDS	38
3.2.1	Noun	38
3.2.1.1	Pronoun	39
3.2.1.2	Proper noun	41
3.2.1.3	Common noun	41
3.2.1.4	Question articles	43
3.2.2	Articles	44
3.2.3	Verb	44
3.3	STRUCTURE OF LEXICON	47
3.3.1	Root	47
3.3.2	Compound Words.....	50
3.3.3	Irregular Verb	51
3.4	ANALYSIS FOR ARABIC WORD	51
3.4.1	Noun	52
3.4.2	Verb	55
3.5	THE MODEL OF ARABIC MORPHOLOGY	55
3.5.1	Classification for Arabic Root	58
3.5.2	The Morphological Analyzer as a Production System.....	62

3.5.2.1 Global database.....	63
3.5.2.2 Rules.....	68
3.5.2.3 Control strategy.....	80
3.5.4 The Output from Morphological Analyzer	81

CHAPTER FOUR ARABIC SYNTAX MODEL

4.1 INTRODUCTION	87
4.2 THE ARABIC SENTENCE STRUCTURE.....	87
4.2.1 Functions Filled by The Arabic Sentence.....	88
4.2.2 The Mood System of The Arabic Sentence.....	88
4.2.3 Prototypical Arabic Sentence Structure.....	92
4.2.3.1 Modifiers.....	92
4.2.3.2 Core sentence functions.....	95
4.3 SYNTACTIC MODEL FOR ARABIC SENTENCE USING ATN	96
4.3.1 The Sentence Network (SEN).....	96
4.3.1.1 Nominal sentence.....	96
4.3.1.2 Verbal sentence	101
4.3.2 Noun Phrase Network (NP).....	109
4.3.2.1 The head of the NP.....	109
4.3.2.2 Modifiers of NP.....	109
4.4 PARSING ALGORITHM	114
4.4.1 Sharing Structure	114
4.4.2 The Form of Resulting Structures	116
4.4.3 Multiple Dimension	120
4.4.4 'Earab' Marks.....	120

CHAPTER FIVE FURTHER SYNTACTIC ISSUES

5.1 INTRODUCTION	122
5.2 INTERROGATIVE SENTENCE	122
5.3 RELATIVE SENTENCE	125
5.4 BOUND SENTENCE	126
5.5 CONJUNCTION	127
5.6 PRONOUN REFERENCE	129

CHAPTER SIX CONCLUSION AND FURTHER WORK

6.1 CONCLUSION	132
6.2 FURTHER WORK.....	134

APPENDIX : A	STRUCTURE OF ARABIC LEXICON
APPENDIX : B	ROOT ENTRIES
APPENDIX : C	PREFIX FOR ARABIC WORD
APPENDIX : D	SUFFIX FOR ARABIC WORD
APPENDIX : E	FORM TABLES
APPENDIX : F	ARABIC SYNTAX MODEL USING ATN

REFERENCE

LIST OF RESERVED WORDS

ENGLISH	ARABIC
Noun	الاسم
Pronoun	الضمير
Proper noun	اسم العلم
Common noun	الاسم العام
Question articles	ادوات الاستفهام
Personal pronoun	الضمائر الشخصية
Gender	الجنس
Person	الدلالة "مخاطب-متكلم-مناقب"
Number	العدد
Connection	الاتصال "متصل- منفصل "
Demonstrative pronoun	اسماء الاشارة
Reference	الاشارة " للقريب-للبعيد "
Relative pronoun	الاسماء الموصولة
Reasonability	العقل
Definition	التعريف
Derivation	الاشتقاق
'Earab' state	حالة الاعراب
'Earab' mark	علامة الاعراب
Articles	الحروف
Transitivity	النعدى
Construction	البناء
Tense	الزمن
'Modareh'	مضارع
Emphasis	التوكيد

ENGLISH

ARABIC

Present-participle	اسم الفاعل
Past-participle	اسم المفعول
'Segha-mobalgha'	صيغة مبالغة
Infinitive	المصدر
Comparative names	اسماء التفضيل
Adjective	الصفة المشبهة
Machine name	اسم الآلة
Form	الوزن
Abstract form	الوزن المجرد
Augmented form	الوزن المزيّد
Weak article	حرف علة
Triliteral salem root	الجذر الثلاثي السالم
Triliteral doubled root	الجذر الثلاثي المضعف
Triliteral weak root	الجذر الثلاثي الناقص
Triliteral hollow root	الجذر الثلاثي الاجوف
Triliteral 'mahmouz' root	الجذر الثلاثي المهموز
Predicate	الخبر
Averb	حال
Classifier	مضاف اليه
Bound sentence	جملة شرط
Object	مفعول به
Declarative sentence	الجملة الخبرية
Interrogative sentence	الجملة الاستفهامية
Relative sentence	جملة صلة
Imperative sentence	جملة الامر

ENGLISH	ARABIC
'Mobtada'	المبتداء
Subject	الفاعل
'Nasikh'	ناسخ
Beginning article	حرف ابتداء
Emphasizing object	مفعول مطلق
Causing Object	مفعول لاجله
Adverbial object	مفعول فيه
Vice-subject	نائب الفاعل
Causing adjective	صفة سببية
Distinction	التمييز
Exception	الاستثناء

LIST OF TABLES

TABLE#	Title	Page
2.1	Comparison between grammars	28
4.1	Articles proceed the present verb	104