

DE LA SALLE UNIVERSITY

**GRACIAS: GRAMMAR ANALYSIS USING
AUGMENTED TRANSITION NETWORKS**

**A Thesis
Presented to the Faculty of the
Graduate School of Education, Arts and Sciences
De La Salle University**

**In Partial Fulfillment
of the Requirements for the Degree of
Master of Science in Teaching
Major in Computer Science**

by

MANUEL MIGUEL MERCADO MERCADO

**Prof. Nopporn Luangprasert
Adviser**

April 1990



EXECUTIVE SUMMARY

Grammar analysis is one important application of natural language processing (NLP), a field in artificial intelligence that is concerned with enabling man to communicate directly with the computer using his own natural language. The main objective addressed by this study is the development of a processor for the English language with grammar-checking capabilities. The study involved analyzing the basic syntax and semantic structures of the English language and representing these using a suitable notation.

An integrated grammar-checking system has been developed, using natural language processing concepts. Aside from the grammar analysing facility, this system includes a built-in text editor, file management routines and dictionary maintenance facilities. The syntax notations were represented using augmented transition networks (ATNs), an NLP processing approach based on state-transition diagrams.

It is recommended that the sentence constructs that may be derived as outputs of the system may be used as inputs to other NLP systems that perform more complicated tasks, like language translation.



DE LA SALLE UNIVERSITY

TABLE OF CONTENTS

EXECUTIVE SUMMARY

ACKNOWLEDGEMENTS

I. INTRODUCTION

1.1	BACKGROUND OF THE STUDY	1
1.2	STATEMENT OF THE RESEARCH PROBLEM	6
1.3	OBJECTIVES OF THE STUDY	8
1.3.1	GENERAL OBJECTIVE	
1.3.2	SPECIFIC OBJECTIVES	
1.4	SIGNIFICANCE OF THE STUDY	9
1.5	METHODOLOGY OF THE STUDY	12
1.6	SCOPE AND LIMITATIONS	16

II. REVIEW OF RELATED LITERATURE 19

2.1	SAD - SAM	21
2.2	SIR	23
2.3	ELIZA	24
2.4	WILKES' MACHINE TRANSLATING SYSTEM	26
2.5	LUNAR	28
2.6	SHRDLU	30
2.7	MARGIE	32
2.8	SAM and PAM	36



2.9	LIFER	39
2.10	NLP: Syntactic Analysis using Semantic Nets	41

III. THEORETICAL FRAMEWORK

3.1	DEFINITION OF TERMS	43
3.2	NATURAL LANGUAGE PROCESSING SYSTEMS	
3.2.1	COMPONENTS OF AN NLP SYSTEM	46
3.2.1.1	PARSER	47
3.2.1.2	LEXICON	48
3.2.1.3	KNOWLEDGE BASE	49
3.2.1.4	OUTPUT GENERATOR	50
3.2.2	APPLICATIONS OF NATURAL LANGUAGE PROCESSING	
3.2.2.1	MACHINE TRANSLATION	51
3.2.2.2	DOCUMENT UNDERSTANDING	53
3.2.2.3	DOCUMENT GENERATION	55
3.2.2.4	NATURAL LANGUAGE FRONT-END	56
3.2.2.5	GRAMMAR ANALYSIS	59
3.3	GRAMMAR	62
3.3.1	FORMAL GRAMMAR	65
3.3.2	TRANSFORMATIONAL GRAMMAR	73
3.3.3	SYSTEMIC GRAMMAR	78
3.3.4	CASE GRAMMAR	83



3.4	PARSING	89
3.4.1	PARSING STRATEGIES	
3.4.1.1	BACKTRACKING vs PARALLEL PROCESSING	90
3.4.1.2	TOP-DOWN vs BOTTOM-UP	91
3.4.2	PARSING ENGLISH SENTENCES	93
3.5	SEMANTIC NETWORKS	72
3.6	TRANSITION NETWORKS	76
3.6.1	RECURSIVE TRANSITION NETWORKS	78
3.6.2	AUGMENTED TRANSITION NETWORKS	82
3.7	CONCEPTUAL DEPENDENCY THEORY	85
3.8	DICTIONARY STRUCTURE	116
3.8.1	SEQUENTIAL FILES	116
3.8.2	DIRECT FILES	118
3.8.2.1	HASHING FUNCTIONS	120
3.8.2.2	COLLISION RESOLUTION TECHNIQUES	121
3.8.3	INDEXED-SEQUENTIAL FILES	123
3.8.4	TREE STRUCTURES	124
3.8.5	B-TREE	126
3.8.6	TRIES	128



3.9	BASIC LINGUISTICS FOR ENGLISH	130
3.9.1	PARTS OF SPEECH	131
3.9.2	CLASSIFICATION OF SENTENCES	134
3.9.3	MORPHOLOGICAL ANALYSIS	137
IV DESIGN ISSUES		
4.1	APPROACHES TO NATURAL LANGUAGE PROCESSING	140
4.1.1	PATTERN-MATCHING APPROACH	141
4.1.2	PHRASE-STRUCTURE GRAMMAR APPROACH	143
4.1.3	SEMANTIC APPROACH	145
4.1.4	KNOWLEDGE-BASED APPROACH	148
4.1.5	NEURAL NETWORK APPROACH	150
4.1.6	TRANSITION NETWORK APPROACH	152
4.1.7	EXTENDED TRANSITION NETWORK APPROACH	154
4.2	DESIGN OF THE ATNs	
4.2.1	GRAMMAR DEFINITION	158
4.2.2	CONSTRUCTION OF ATN DIAGRAMS	165
4.2.3	ALGORITHM DESIGN FOR ATNs	167
4.3	DESIGN OF MAJOR DATA STRUCTURES	
4.3.1	LEXICON	176
4.3.2	STRUCTURE FOR TEMPORARY WORDS	178
4.3.3	STRUCTURE FOR INPUT SENTENCE	181



V.	IMPLEMENTATION ISSUES	
5.1	SYSTEM OVERVIEW	182
5.2	SCOPE AND LIMITATIONS OF THE SYSTEM	183
5.3	SOFTWARE SUBSYSTEMS	
5.3.1	FILE MANAGEMENT	185
5.3.2	TEXT EDITOR	189
5.3.3	GRAMMAR CHECKER	189
5.3.4	DICTIONARY MAINTENANCE	194
5.4	IMPLEMENTATION OF LEXICON and WORD STRUCTURES	
5.4.1	LEXICON OR DICTIONARY FILE	197
5.4.2	B-TREE STRUCTURE for INDEX TABLE	199
5.4.3	TEMPORARY WORDS STRUCTURE	202
5.4.4	INPUT-SENTENCE STRUCTURE	204
5.5	IMPLEMENTATION OF GRAMMAR ANALYZER	
5.5.1	LEXICAL ANALYSIS	206
5.5.1.1	LEXICAL VALIDATION	207
5.5.1.2	DIRECTORY LOOK UP	209
5.5.1.3	MORPHOLOGICAL ANALYSIS	211
5.5.1.4	RESOLVE AMBIGUITIES	212
5.5.2	SYNTACTIC ANALYSIS	215
5.5.3	SEMANTIC ANALYSIS	220



DE LA SALLE UNIVERSITY

x

VI. CONCLUSIONS AND FUTURE DIRECTIONS

6.1 CONCLUSIONS	229
6.2 FUTURE DIRECTIONS	231

BIBLIOGRAPHY	233
------------------------	-----

APPENDICES

APPENDIX A. PARTS OF SPEECH	238
APPENDIX B: NOTATION CODES	241
APPENDIX C: TERMINAL SYMBOLS	245
APPENDIX D: SUBSET GRAMMAR and ATNs	251

