CONTENTS

LIST OF FIGURES xi

LIST OF TABLES xv

PREFACE xvii

1. INTRODUCTION 1

1.1. MOTIVATION 1
1.2. SCOPE OF THE BOOK 3
1.3. CHARACTERISTICS OF NATURAL LANGUAGE 5
1.4. PREVIOUS WORK 7

2. INFORMAL DESCRIPTION OF THE ALGORITHM 9

2.1. INTRODUCTION 9
2.2. THE LR PARSING ALGORITHM 9
2.2.1. AN EXAMPLE 9
2.2.2. PROBLEM IN APPLYING TO NATURAL LANGUAGES 12
2.3. HANDLING MULTIPLE ENTRIES 13
2.3.1. WITH STACK LIST 14
2.3.2. WITH A TREE-STRUCTURED STACK 15
2.3.3. WITH A GRAPH-STRUCTURED STACK 16
2.4. AN EFFICIENT REPRESENTATION OF A PARSE FOREST 17
2.4.1. SUB-TREE SHARING 18
2.4.2. LOCAL AMBIGUITY PACKING 19

3. EXAMPLES 21

3.1. INTRODUCTION 21
3.2. THE EXAMPLE 22
3.3. MANAGING MULTI-PART-OF-SPEECH WORDS 38
3.4. MANAGING UNKNOWN WORDS 43

4. FORMAL SPECIFICATION OF THE ALGORITHM 47

4.1. INTRODUCTION 47
4.2. THE RECOGNIZER 47
4.2.1. TERMS IN GRAPH THEORY 48
4.2.2. THE GRAPH-STRUCTURED STACK 48
4.2.3. THE ALGORITHM FOR LR GRAMMARS 49
4.2.4. THE ALGORITHM FOR NON-E GRAMMARS 52
4.2.5. THE ALGORITHM FOR E-GRAMMARS 56
4.2.6. THE FULL VERSION OF THE ALGORITHM 60
4.3. THE PARSER 64
4.3.1. ADDITIONAL TERMS IN GRAPH THEORY 64

4.3.2. THE PARSE FOREST 64
4.3.3. THE PARSING ALGORITHM 65
4.4. SUMMARY 69

5. COMPARISON WITH OTHER ALGORITHMS 71

5.1. INTRODUCTION 71
5.2. RECOGNITION TIME 72
5.3. PARSE FOREST REPRESENTATION 73
5.4. THE DEFECT OF EARLEY'S FOREST REPRESENTATION 74
5.5. SUMMARY 77

6. EMPIRICAL RESULTS 79

6.1. INTRODUCTION 79
6.1.1. THE PROGRAMS 79
6.1.2. THE SAMPLE GRAMMARS 81
6.1.3. THE SAMPLE SENTENCES 81
6.2. PARSING TIME 82
6.2.1. WITH RESPECT TO SENTENCE LENGTH 82
6.2.2. WITH RESPECT TO SENTENCE AMBIGUITY 84
6.2.3. WITH RESPECT TO GRAMMAR SIZE 84
6.3. COMPARISON WITH EARLEY'S ALGORITHMS 86
6.3.1. WITH RESPECT TO SENTENCE LENGTH 86
6.3.2. WITH RESPECT TO SENTENCE AMBIGUITY 86
6.3.3. WITH RESPECT TO GRAMMAR SIZE 91
6.4. SPACE EFFICIENCY 91
6.4.1. PARSE FOREST REPRESENTATION 91
6.4.2. GRAPH-STRUCTURED STACK 91
6.4.3. SIZE OF PARSING TABLES 94
6.5. SUMMARY 94

7. LEFT-TO-RIGHT ON-LINE PARSING 95

7.1. INTRODUCTION 95
7.2. RESPONSE TIME 95
7.3. UNDO-ABILITY 96
7.4. EARLY ERROR DETECTION 97
7.5. ON-LINE PARSING AND ROBUST PARSING 98
7.5.1. LEXICAL LEVEL EXTRAGRAMMATICALITIES 99
7.5.2. SENTENTIAL LEVEL EXTRAGRAMMATICALITIES 100
7.5.3. DIALOG LEVEL EXTRAGRAMMATICALITIES 101
7.6. SUMMARY 101

8. SENTENCE DISAMBIGUATION BY ASKING 103

8.1. INTRODUCTION 103
8.2. EXPLANATION LIST COMPARISON 104
8.2.1. THE REVISED VERSION OF ELC 105
8.2.2. HEAD 106
8.2.3. MULTIPLE EXPLANATIONS 107

8.2.4. REVISED ELC 107
8.2.5. MORE COMPLEX EXAMPLE 109
8.2.6. SIMPLE QLIST CONSTRUCTION ALGORITHM 111
8.3. DISAMBIGUATION OUT OF THE SHARED-PACKED FOREST 112
8.3.1. FOCUS 112
8.3.2. FOCUSED QLISTS 112
8.3.3. SHAVING A FOREST 114
8.4. SUMMARY 117

9. INTERACTIVE/PERSONAL MACHINE TRANSLATION 121

9.1. INTRODUCTION 121
9.2. THE DESIGN PHILOSOPHY 121
9.3. THE DESIGN DECISIONS 124
9.4. THE EXPERIMENTAL INTERACTIVE SYSTEM 124
9.5. FOUR KINDS OF AMBIGUITIES 126
9.5.1. AMBIGUITY BY MULTI-PART-OF-SPEECH WORDS [TYPE A] 126
9.5.2. AMBIGUITY BY CONJUNCTION [TYPE B] 127
9.5.3. AMBIGUITY BY MODIFICATION [TYPE C] 127
9.5.4. AMBIGUITY BY GENERAL AND SPECIFIC RULES [TYPE D] 127
9.6. EMPIRICAL RESULTS 128
9.7. DISCUSSIONS 130
9.8. SUMMARY 131

10. CONCLUDING REMARKS 133

10.1. SUMMARY OF THE BOOK 133
10.2. FUTURE WORK 134

APPENDIX A. THE PARSING TABLE CONSTRUCTOR 135

APPENDIX B. EARLEY'S ALGORITHM 139

APPENDIX C. PROOF OF CORRECTNESS OF THE ALGORITHM 141

C.1. INTRODUCTION 141
C.2. SOUNDNESS OF THE ALGORITHM 141
C.3. COMPLETENESS OF THE ALGORITHM 146

APPENDIX D. RAW EMPIRICAL DATA 151

APPENDIX E. PROGRAMS USED IN THE EXPERIMENTS 155

E.1. TOMITA'S ALGORITHM 155
E.2. EARLEY'S ALGORITHM 160
E.3. EARLEY'S ALGORITHM WITH AN IMPROVEMENT 162
E.4. LR(0) TABLE CONSTRUCTION ALGORITHM 166
E.5. UTILITY FUNCTIONS 169

APPENDIX F. GRAMMARS USED IN THE EXPERIMENTS 171

APPENDIX G. SENTENCES USED IN THE EXPERIMENTS 185
APPENDIX H. NISHIDA AND DOSHITA'S SYSTEM 191
REFERENCES 193
SUBJECT INDEX 199
AUTHOR INDEX 201

List of Figures

Figure 1-1: Dense Ambiguity 6
Figure 1-2: Infinite Ambiguity 6
Figure 1-3: Cyclic Grammar 6
Figure 2-1: Example Grammar 10
Figure 2-2: LR Parsing Table 10
Figure 2-3: Trace of LR Parsing 11
Figure 2-4: Context-free Grammars and LR grammars 12
Figure 2-5: An Example Ambiguous Grammar 12
Figure 2-6: LR Parsing Table with Multiple Entries 13
Figure 2-7: Stack List 15
Figure 2-8: A Tree-structured Stack 16
Figure 2-9: A Graph-Structured Stack 17
Figure 2-10: Shared Forest 18
Figure 2-11: Unpacked Shared Forest 20
Figure 2-12: Packed Shared Forest 20
Figure 3-1: Trace of the Parser 23
Figure 3-2: Trace of the Parser (cont.) 23
Figure 3-3: Trace of the Parser (cont.) 23
Figure 3-4: Trace of the Parser (cont.) 24
Figure 3-5: Trace of the Parser (cont.) 24
Figure 3-6: Trace of the Parser (cont.) 24
Figure 3-7: Trace of the Parser (cont.) 25
Figure 3-8: Trace of the Parser (cont.) 25
Figure 3-9: Trace of the Parser (cont.) 25
Figure 3-10: Trace of the Parser (cont.) 26
Figure 3-11: Trace of the Parser (cont.) 26
Figure 3-12: Trace of the Parser (cont.) 27
Figure 3-13: Trace of the Parser (cont.) 27
Figure 3-14: Trace of the Parser (cont.) 28
Figure 3-15: Trace of the Parser (cont.) 28
Figure 3-16: Trace of the Parser (cont.) 29
Figure 3-17: Trace of the Parser (cont.) 29
Figure 3-18: Trace of the Parser (cont.) 30
Figure 3-19: Trace of the Parser (cont.) 30
Figure 3-20: Trace of the Parser (cont.) 31
Figure 3-21: Trace of the Parser (cont.) 31
Figure 3-22: Trace of the Parser (cont.) 32
Figure 3-23: Trace of the Parser (cont.) 32

Figure 3-25: Trace of the Parser (cont.) 33
Figure 3-26: Trace of the Parser (cont.) 34
Figure 3-27: Trace of the Parser (cont.) 34
Figure 3-28: Trace of the Parser (cont.) 35
Figure 3-29: Trace of the Parser (cont.) 35
Figure 3-30: Trace of the Parser (cont.) 36
Figure 3-31: Trace of the Parser (cont.) 36
Figure 3-32: Trace of the Parser (final) 37
Figure 3-33: Another Example Grammar 38
Figure 3-34: Parsing Table for figure 3-33 38
Figure 3-35: Trace of the Parser 38
Figure 3-36: Trace of the Parser (cont.) 39
Figure 3-37: Trace of the Parser (cont.) 39
Figure 3-38: Trace of the Parser (cont.) 39
Figure 3-39: Trace of the Parser (cont.) 40
Figure 3-40: Trace of the Parser (cont.) 40
Figure 3-41: Trace of the Parser (cont.) 40
Figure 3-42: Trace of the Parser (cont.) 41
Figure 3-43: Trace of the Parser (cont.) 41
Figure 3-44: Trace of the Parser (cont.) 41
Figure 3-45: Trace of the Parser (final) 42
Figure 3-46: Trace of the Parser 43
Figure 3-47: Trace of the Parser (cont.) 43
Figure 3-48: Trace of the Parser (cont.) 43
Figure 3-49: Trace of the Parser (cont.) 44
Figure 3-50: Trace of the Parser (cont.) 44
Figure 3-51: Trace of the Parser (cont.) 44
Figure 3-52: Trace of the Parser (cont.) 45
Figure 3-53: Trace of the Parser (cont.) 45
Figure 3-54: Trace of the Parser (final) 45
Figure 4-1: Example of REDUCER 52
Figure 4-2: Example of SHIFTER 52
Figure 4-3: The Graph-Structured Stack 53
Figure 4-4: Example before Refinement 60
Figure 4-5: Example after Refinement 60
Figure 4-6: Another Example before Refinement 61
Figure 4-7: Another Example after Refinement 61
Figure 5-1: Heavy Ambiguity 72
Figure 5-2: Infinite Ambiguity 73
Figure 5-3: Cyclic Grammar 73
Figure 5-4: Grammar G1 74
Figure 5-5: Earley's Parse Forest 74
Figure 5-6: Underlying Parse Trees 75
Figure 5-7: Grammar G2 75
Figure 5-8: Correct Parse Trees 75
Figure 5-9: Defective Representation 75

Figure 5-10: Wrong Parse Trees 76
Figure 6-1: Time against Sentence Length 83
Figure 6-2: Time against Sentence Length 83
Figure 6-3: Time against Ambiguity (a) 84
Figure 6-4: Time against Ambiguity (b) 85
Figure 6-5: Time against Grammar Size 85
Figure 6-6: Earley's against Tomita's 87
Figure 6-7: Earley/Tomita ratio against Sentence Length (a) 88
Figure 6-8: Earley/Tomita ratio against Sentence Length (b) 88
Figure 6-9: Earley/Tomita ratio against Sentence Ambiguity (a) 89
Figure 6-10: Earley/Tomita ratio against Sentence Ambiguity (b) 89
Figure 6-11: Earley/Tomita ratio against Sentence Ambiguity (c) 90
Figure 6-12: Earley/Tomita ratio against Sentence Ambiguity (d) 90
Figure 6-13: Earley/Tomita ratio against Grammar Size (a) 92
Figure 6-14: Earley/Tomita ratio against Grammar Size (b) 92
Figure 6-15: Size of Parse Forest against Sentence Ambiguity 93
Figure 6-16: Size of Graph-structured Stack against Sentence Ambiguity 93
Figure 6-17: Size of Grammar and Size of its Parsing Tables 94
Figure 7-1: Example of Online Parsing 96
Figure 7-2: Example of Undoing 97
Figure 7-3: Example of Earley-Error Detection 98
Figure 7-4: Example of Online Spelling Correction 100
Figure 8-1: Example Grammar and Explanation Templates with a Focus 113
Figure 8-2: Parse Forest with Explanations 113
Figure 8-3: A Set of focused Qlists and pointers 114
Figure 8-4: Forest before shaving off 116
Figure 8-5: Forest after shaving off node 13 117
Figure 8-6: Forest after shaving off subnode 22-1 118
Figure 8-7: Final forest after disambiguation 119

List of Tables

Table 4-1: Pre-defined functions and global variables 50
Table 4-2: Pre-defined functions and global variables 54
Table 4-3: Global variables 56
Table 4-4: Pre-defined functions and global variables 65
Table 6-1: Ambiguity of Sentence Set II 82
Table 9-1: Sentence Length and Number of Questions 129
Table D-1: Parsing Time on Sentence Set I 152
Table D-2: Parsing Time on Sentence Set II 153
Table D-3: Parsing Time on Sentence Set II (cont.) 153
Table D-4: Space Efficiency of Program I 154