Natural Language Processing for Prolog Programmers

Michael A. Covington

Artificial Intelligence Programs
The University of Georgia
Athens, Georgia

PRENTICE HALL, Upper Saddle River, New Jersey 07458
Contents

PREFACE xv

1 NATURAL LANGUAGE 1

1.1 What is NLP? 1
1.2 Language from a scientific viewpoint 2
1.3 Language and the brain 4
1.4 Levels of linguistic analysis 5
  1.4.1 Phonology, 5
  1.4.2 Morphology, 6
  1.4.3 Syntax, 7
  1.4.4 Semantics, 9
  1.4.5 Pragmatics, 10
1.5 Why use Prolog? 12
1.6 Further reading 12

2 TEMPLATES AND KEYWORDS 14

2.1 Template matching 14
2.1.1 ELIZA, 14
2.1.2 Other template systems, 15

2.2 DOS commands in English 16
2.2.1 Recipe for a template system, 16
2.2.2 Implementing simplification rules, 19
2.2.3 Implementing translation rules, 20
2.2.4 The rest of the system, 21

2.3 Keyword analysis 24
2.3.1 A time-honored approach, 24
2.3.2 Database querying, 25
2.3.3 A database in Prolog, 26
2.3.4 Building a keyword system, 27
2.3.5 Constructing a query, 28
2.3.6 Lambda abstraction, 30
2.3.7 Lambdas in Prolog, 31
2.3.8 The complete system, 32

2.4 Toward more natural input 34
2.4.1 Ellipsis, 34
2.4.2 Anaphora, 34

3 DEFINITE-CLAUSE GRAMMARS 36

3.1 Phase structure 36
3.1.1 Trees and PS rules, 36
3.1.2 Phase-structure formalism, 38
3.1.3 Recursion, 40

3.2 Top-down parsing 42
3.2.1 A parsing algorithm, 42
3.2.2 Parsing with Prolog rules, 43

3.3 DCG rules 45
3.3.1 DCG notation, 45
3.3.2 Loops, 48
3.3.3 Some details of implementation, 50

3.4 Using DCG parsers 51
3.4.1 Building syntactic trees, 51
3.4.2 Agreement, 54
3.4.3 Case marking, 56
3.4.4 Subcategorization, 57
3.4.5 Undoing syntactic movements, 59
3.4.6 Separating lexicon from PS rules, 61
3.5 Building semantic representations 63
3.5.1 Semantic composition, 63
3.5.2 Semantic grammars, 66
3.6 Offbeat uses for DCG rules 66
3.7 Excursus: transition-network parsers 68
3.7.1 States and transitions, 68
3.7.2 Recursive transition networks, 71
3.7.3 Augmented transition networks (ATNs), 72

4 ENGLISH PHRASE STRUCTURE 77

4.1 Phrase structure 77
4.1.1 Trees revisited, 78
4.1.2 Constituents and categories, 79
4.1.3 Structural ambiguity, 80
4.2 Traditional grammar 82
4.2.1 Parts of speech, 82
4.2.2 Grammatical relations, 84
4.3 The noun phrase and its modifiers 85
4.3.1 Simple NPs, 85
4.3.2 Multiple adjective positions, 87
4.3.3 Adjective phrases, 88
4.3.4 Sentences within NPs, 88
4.4 The verb phrase 90
4.4.1 Verbs and their complements, 90
4.4.2 Particles, 93
4.4.3 The copula, 93
4.5 Other structures 96
4.5.1 Conjunctions, 96
4.5.2 Sentential PPs, 99
4.6 Where PS rules fail 101
4.6.1 Adverbs and ID/LP formalism, 101
4.6.2 Postposing of long constituents, 104
4.6.3 Unbounded movements, 105
4.6.4 Transformational grammar, 108
4.7 Further reading 110

5 UNIFICATION-BASED GRAMMAR 111

5.1 A unification-based formalism 111
5.1.1 The problem, 111
5.1.2 What is UBG?, 111
5.1.3 How features behave, 112
5.1.4 Features and PS rules, 114
5.1.5 Feature-structure unification, 115

5.2 A sample grammar 117
5.2.1 Overview, 117
5.2.2 Lexical entries, 118
5.2.3 Phrase-structure rules, 119
5.2.4 How the rules fit together, 121

5.3 Formal properties of feature structures 123
5.3.1 Features and values, 123
5.3.2 Re-entrancy, 126
5.3.3 Functions, paths, and equational style, 128

5.4 An extension of Prolog for UBG 130
5.4.1 A better syntax for features, 130
5.4.2 Translating a single feature structure, 132
5.4.3 Translating terms of all types, 134
5.4.4 Translating while consulting, 136
5.4.5 Output of feature structures, 138

5.5 UBG in theory and practice 141
5.5.1 A more complex grammar, 141
5.5.2 Context-free backbones and subcategorization lists, 146
5.5.3 Negative and disjunctive features, 149

6 PARSING ALGORITHMS 151

6.1 Comparing parsing algorithms 151
6.2 Top-down parsing 151
6.3 Bottom-up parsing 155
6.3.1 The shift-reduce algorithm, 155
6.3.2 Shift-reduce in Prolog, 157
6.4 Left-corner parsing 158
6.4.1 The key idea, 158
6.4.2 The algorithm, 160
6.4.3 Links, 162
6.4.4 BUP, 165
6.5 Chart parsing 167
Contents

6.5.1 The key idea, 167
6.5.2 A first implementation, 168
6.5.3 Representing positions numerically, 170
6.5.4 Completeness, 171
6.5.5 Subsumption, 174

6.6 Earley’s algorithm 176
6.6.1 The key idea, 176
6.6.2 An implementation, 177
6.6.3 Predictor, 180
6.6.4 Scanner, 182
6.6.5 Completer, 182
6.6.6 How Earley’s algorithm avoids loops, 183
6.6.7 Handling null constituents, 184
6.6.8 Subsumption revisited, 185
6.6.9 Restriction, 186
6.6.10 Improving Earley’s algorithm, 188
6.6.11 Earley’s algorithm as an inference engine, 188

6.7 Which parsing algorithm is really best? 191
6.7.1 Disappointing news about performance, 191
6.7.2 Complexity of parsing, 193
6.7.3 Further reading, 195

7 SEMANTICS, LOGIC, AND MODEL THEORY 196

7.1 The problem of semantics 196
7.2 From English to logical formulas 197
7.2.1 Logic and model theory, 197
7.2.2 Simple words and phrases, 198
7.2.3 Semantics of the $N_1$ constituent, 201
7.3 Quantifiers (determiners) 203
7.3.1 Quantifiers in language, logic, and Prolog, 203
7.3.2 Restrictor and scope, 205
7.3.3 Structural importance of determiners, 207
7.3.4 Building quantified structures, 207
7.3.5 Scope ambiguities, 212
7.4 Question answering 214
7.4.1 Simple yes/no question, 214
7.4.2 Getting a list of solutions, 215
7.4.3 Who/what/which questions, 217
7.5 From formula to knowledge base 219
7.5.1 Discourse referents, 219
7.5.2 Anaphora, 221
7.5.3 Definite reference (the), 224
7.5.4 Plurals, 225
7.5.5 Mass nouns, 227

7.6 Negation 228
7.6.1 Negative knowledge, 228
7.6.2 Negation as a quantifier, 229
7.6.3 Some logical equivalences, 230

7.7 Further reading 232

8 FURTHER TOPICS IN SEMANTICS 233

8.1 Beyond model theory 233
8.2 Language translation 233
8.2.1 Background, 233
8.2.2 A simple technique, 234
8.2.3 Some Latin grammar, 235
8.2.4 A working translator, 237
8.2.5 Why translation is hard, 238

8.3 Word-sense disambiguation 239
8.3.1 The problem, 239
8.3.2 Disambiguation by activating contexts, 240
8.3.3 Finding the best compromise, 244
8.3.4 Spreading activation, 247

8.4 Understanding events 248
8.4.1 Event semantics, 248
8.4.2 Time and tense, 251
8.4.3 Scripts, 252

8.5 Further reading 256

9 MORPHOLOGY AND THE LEXICON 257

9.1 How morphology works 257
9.1.1 The nature of morphology, 257
9.1.2 Morphemes and allomorphs, 258

9.2 English inflection 260
9.2.1 The system, 269
9.2.2 Morphographemics (spelling rules), 262

9.3 Implementing English inflection 263