

Contents

List of Figures	viii
List of Tables	xiii
Chapter 1 Introduction	1
1.1 Basic Assumptions	4
1.2 Overview of the Research Results	5
1.2.1 Representation of Natural Language	6
1.2.2 Grammar Formalism	11
1.2.3 Learning Model	16
1.2.4 Applications of the LWFG Learning Model	19
1.3 Contributions	25
1.4 Dissertation Guide	28
Chapter 2 Related Work	31
2.1 Grammar Formalism	32
2.2 Natural Language Representation	35
2.3 Grammar Learning	36

I	Theoretical Part	39
Chapter 3	Lexicalized Well-Founded Grammars	40
3.1	Well-Founded Grammars	41
3.2	Augmenting Well-Founded Grammars with Semantics	44
3.2.1	Semantic Molecule	45
3.2.2	Semantic Composition and Semantic Interpretation as Grammar Constraints	49
3.2.3	Lexicalized Well-Founded Grammars: Definition	51
3.3	Derivation in LWFGs	52
3.4	Semantics of LWFGs	53
3.5	Parser/Generator for LWFGs	55
3.5.1	Parsing/Generation as Deduction	56
3.5.1.1	Basic Notions	57
3.5.1.2	Robust Bottom-up Active Chart Parser/Generator	58
3.6	Representative Examples of a LWFG	62
Chapter 4	Semantic Composition and Semantic Interpretation as Grammar Constraints	69
4.1	Semantic Composition: Principles and Properties	70
4.1.1	Body Variable Specialization Substitution(ν)	75
4.1.2	Head Feature Generalization Substitution (μ)	76
4.1.3	Compositional Constraint Determinacy	78
4.1.4	Rule Generalization	80
4.1.4.1	Rule Derivation Step	80

4.1.4.2	Rule Generalization Step	82
4.1.5	Example	84
4.2	Ontology-based Semantic Interpretation	85
4.2.1	The Semantic Interpreter	88
Chapter 5 Induction of Lexicalized Well-Founded Grammars		90
5.1	Grammar Learning	90
5.1.1	Properties and Assumptions	91
5.1.2	Relational Learning from Ordered Representative Examples	98
5.1.3	Iterative Learning from Unordered Representative Examples	103
5.2	Grammar Merging	105
5.3	Grammar Approximation by Representative Sublanguage (GARS) .	109
5.3.1	Foundation of the Search Space for Grammar Induction . . .	109
5.3.1.1	Representative Examples Parsing Preserving Gram- mars	110
5.3.1.2	Semantic-based Complete Grammar Lattice	115
5.3.1.3	Grammar Boolean Algebra	121
5.3.1.4	Learnability Theorem	124
5.3.2	Grammar Induction Model	132
5.3.2.1	Grammar Induction Problem in ILP setting	133
5.3.2.2	Decidability of LWFGs	135
5.3.2.3	GARS Model	136

II Applicative Part 139

Chapter 6 Expressiveness of the Semantic Representation (OntoSeR) 140

6.1	The Lexicon and Elementary Semantic Molecules	148
6.1.1	Verbs	149
6.1.2	Nouns	154
6.1.3	Adjectives and Adverbs	157
6.1.4	Prepositions	160
6.1.5	Coordination	162
6.2	Raising and Control	166
6.2.1	Linguistic Phenomenon and Syntactic Analysis	166
6.2.2	Semantic Analysis	170
6.2.2.1	Subject Raising and Subject Control	170
6.2.2.2	Object Raising and Object Control	172
6.2.2.3	Passivizing the Complement of the Matrix Verb	173
6.3	Relative Clauses and Wh-Questions	174
6.3.1	Relative Clauses	176
6.3.2	Wh-Questions	179
6.3.3	Filler-Gap Dependencies	182
6.4	Ambiguity	184

Chapter 7 A General Framework for LWFG Learning 187

7.1	Learning an Experimental LWFG	189
7.1.1	Training Data	190
7.1.2	Overall Results	191

7.1.3	Controlled Experiments	201
7.2	Concluding Remarks	204
Chapter 8 OntoSeR for Knowledge Acquisition		206
8.1	Copula <i>to be</i>	211
8.1.1	Copula <i>be</i> for terminological knowledge	214
8.2	Application: Terminology Acquisition in the Medical Domain	218
8.2.1	DEFINDER: Getting NL Definitions from Corpora	219
8.2.2	Characteristics of the Definitional Corpus	221
8.2.3	Terminological Knowledge Acquisition	224
8.3	Meaning as Answers to Questions	236
8.4	Acquisition and Querying of OKR-annotated treebanks — Pilot Experiment	240
Chapter 9 Conclusions		246
9.1	Contributions	246
9.1.1	Theoretical Contributions	247
9.1.2	Application-oriented Contributions	250
9.2	Open Problems and Future Work	252
9.2.1	Weak Semantic Context/Weak Concept Identity	252
9.2.2	Evaluation	253
Appendix A Learning an Experimental Grammar		255
A.1	Learning Auxiliary Verbs	256
A.1.1	Learning From Ordered Representative Examples	256
A.1.2	Learning from Unordered Representative Examples	261

A.2	Example of Grammar Learning Steps	264
A.3	Examples of Parsing/Generation	267
A.3.1	Finite Verbs	267
A.3.2	Relative Clause (without agreement)	268
A.3.3	Relative Clause (with agreement)	269
A.3.4	Complex Utterance	270
A.4	Samples of the Learned Grammar: Noun Compounds, Raising/Control, Reduced Relative Clauses	271
A.4.1	Representative Examples	272
A.4.2	Learned Grammar Rules and Samples of Compositional Con- straints	273
A.5	Long-Distance Dependencies in Wh-questions	278
A.6	Generating Representative Examples	279
Appendix B Example of an Ambiguous Utterance		283
Appendix C Acquisition and Querying of a Pilot OKR-annotated treebank		289
C.1	Medical Definitions	289
C.2	Acquired Pilot OKR-annotated treebank	291
C.3	NL-Querying Experiment	294
C.4	Samples of Full Concept-level Answers to NL-Queries	298
Appendix D DEFINDER Evaluation		304
D.1	Quantitative Evaluation	304
D.2	Qualitative Evaluation: Lay User Perspective	305

D.3 Qualitative Evaluation: Medical Specialist Perspective	307
D.4 Coverage of Existing Dictionaries	308
References	310