

CSLI Studies in
Computational Linguistics

Jacy
An Implemented
Grammar of Japanese

Melanie Siegel, Emily M. Bender,
and Francis Bond

Contents

Preface **vii**

Notational Conventions and Symbols xv

1 Introduction 1

1.1 History of the Grammar 4

1.2 Theoretical Framework and Development Tools 5

1.3 **Jacy** among other HPSG Grammars of Japanese 6

1.4 **Jacy** and Multilingual Grammar Engineering 6

1.5 How this Book is Structured 7

1.6 Further Reading 7

2 Jacy in Different Application Domains 9

2.1 Appointment Scheduling in Machine Translation 10

2.2 Email in the Banking Domain 15

2.3 Hybrid Language Processing 17

2.4 Dictionary Definition Sentences 20

2.5 Open Source Machine Translation 23

2.6 The Full **Jacy** System 27

2.7 Summary and Further Reading 28

3 Basic Japanese Phrase Structure 29

3.1 Overview of HPSG 29

3.2 Core Phrase Structure Rules 30

3.2.1 Head-Subject Phrase 32

3.2.2 Head-Complement Phrases 33

3.2.3	Head-Adjunct Phrases	35
3.2.4	Coordinated Structures	37
3.2.5	Head-Specifier Constructions	41
3.2.6	Head-Marker Constructions	42
3.2.7	Utterance Rules	44
3.3	Head-Initial Constructions	45
3.3.1	The Position of Syntactic Heads in Japanese	45
3.3.2	Head-Initial Modification	49
3.3.3	Head-Initial Complementation	56
3.3.4	Corpus Study	60
3.4	Semantic Representations	61
3.5	Summary and Further Reading	65
4	Word Order and Subcategorization	67
4.1	Properties of Japanese Arguments	67
4.2	Previous Analyses	70
4.3	Jacy's Approach	71
4.3.1	The Feature OPT	71
4.3.2	Required Adjacency	72
4.3.3	Zero Pronominalization and Obligatory Overt Realization	73
4.4	Summary and Further Reading	75
5	Verbs and Adjectives	77
5.1	Verbal Subcategorization Types	77
5.1.1	Intransitive Verbs	78
5.1.2	Transitive Verbs	80
5.1.3	Ditransitive Verbs	83
5.1.4	Copula Verbs	85
5.1.5	Adjective Subcategorization Types	89
5.2	Inflectional and Derivational Rules	89
5.2.1	Stem Classes	91
5.2.2	Ending Types	93
5.2.3	Inflectional Rules	93
5.2.4	Sample Derivation	95
5.2.5	Stand-Alone Stems	98
5.2.6	Derivational Rules	98
5.2.7	Adjective Inflectional Types	99
5.3	Auxiliary Constructions	100
5.3.1	Aspect Auxiliaries	100

5.3.2	Content Auxiliaries	102
5.3.3	Perspective Auxiliaries	103
5.3.4	Auxiliaries and Honorification	104
5.3.5	Contracted Auxiliaries	105
5.3.6	Summary	106
5.4	Passive Constructions	106
5.4.1	Simple Passives	106
5.4.2	Adversative Passive	109
5.4.3	Honorification with Passive Morphology	111
5.4.4	Potential with Passive Morphology	112
5.5	Causative	112
5.6	Summary and Further Reading	116
6	Nominal Structures: Linking Syntax, Semantics and Pragmatics	119
6.1	Ordinary Nouns and Noun Phrases	120
6.2	Names and Named Entities	123
6.3	Pronouns	125
6.3.1	Demonstrative Pronouns	125
6.3.2	Locative Pronouns	126
6.3.3	Personal Pronouns	127
6.3.4	The Reflexive <i>jibun</i>	128
6.4	Nominalizers	134
6.5	Temporal Nouns	136
6.6	Noun Modification	138
6.6.1	Noun Modification by Post-Positional Phrases	138
6.6.2	Relative Clause Constructions	139
6.6.3	Pre-Nominal Adjectives	142
6.7	Numeral Classifiers	143
6.7.1	Semantic Representations	147
6.7.2	The Analysis	149
6.8	Summary and Further Reading	159
7	Particles	163
7.1	Cooccurrence of Particles	164
7.2	The Type Hierarchy of Japanese Particles	168
7.3	Case Particles <i>case-p-lex</i>	169
7.3.1	General Properties of Case Particles	170

7.3.2	The Case Particle <i>ga</i>	175
7.3.3	The Case Particle <i>wo</i>	177
7.3.4	The Case Particle <i>ni</i>	180
7.3.5	Other Case Particles	181
7.4	Other Semantically Empty Particles	182
7.4.1	Complementizers	182
7.4.2	Coordination Particles	183
7.5	Particles with Semantic Content <i>cont-p-lex</i>	184
7.5.1	Adnominal-Specifying <i>no</i>	184
7.5.2	Modifying Particles <i>mod-p-lex</i>	185
7.5.3	Topic Particles	189
7.5.4	Noun Phrase Conjunctions <i>conj-p-lex</i>	196
7.6	Sentence Particles and Sentence Force <i>sa-p-lex</i>	198
7.7	Omitted Particles	199
7.8	Evaluation of Case and Modifying Particles	200
7.9	Summary and Further Reading	201
8	Other Word Classes	203
8.1	Adverbs	203
8.1.1	Adverb Morphology and Derived Adverbs	203
8.1.2	Types of Adverbial Modification	204
8.2	Interrogatives	205
8.3	Demonstratives	206
8.4	Summary and Further Reading	208
9	Honorifics	209
9.1	Honorific Forms in Japanese	210
9.1.1	Three Types of Honorifics	210
9.1.2	Interaction of Different Kinds of Honorification	212
9.2	Previous Approaches	214
9.3	Japanese Honorification in Jacy	217
9.4	Evaluation	223
9.5	Honorification and Machine Translation	223
9.6	Honorification in Other Languages	224
9.7	Summary	225

10	Grammar Engineering	227
10.1	The Development Environment	227
10.2	The Grammar Development Cycle	228
10.3	Integrating Treebanking	230
10.4	Embedded Documentation: The Lexical Type Database	234
10.4.1	Linguistic Discussion	235
10.4.2	Exemplification	236
10.4.3	Implementation	236
10.4.4	Building the Lexical Type Database	238
10.5	Automatic Error Detection and Correction	239
10.6	Summary and Further Reading	239
11	The Current State of the Grammar	243
11.1	What is the Size of the Grammar?	244
11.2	What is the General Coverage on Different Kinds of Data?	245
11.2.1	Test Suites and Coverage	246
11.2.2	The Hinoki Treebank	247
11.3	How Can We Select the Best Interpretation?	250
11.4	How Domain-Adaptable is the Grammar?	252
11.5	How Far Can the Grammar Be Used in Multilingual Applications?	254
11.6	Summary and Further Reading	254
12	Conclusion	255
	References	261
	Index	279