

Contents

Foreword	23
Preface	25
About the Authors	31
1 Introduction	35
1.1 Knowledge in Speech and Language Processing	36
1.2 Ambiguity	38
1.3 Models and Algorithms	39
1.4 Language, Thought, and Understanding	40
1.5 The State of the Art	42
1.6 Some Brief History	43
1.6.1 Foundational Insights: 1940s and 1950s	43
1.6.2 The Two Camps: 1957–1970	44
1.6.3 Four Paradigms: 1970–1983	45
1.6.4 Empiricism and Finite-State Models Redux: 1983–1993	46
1.6.5 The Field Comes Together: 1994–1999	46
1.6.6 The Rise of Machine Learning: 2000–2008	46
1.6.7 On Multiple Discoveries	47
1.6.8 A Final Brief Note on Psychology	48
1.7 Summary	48
Bibliographical and Historical Notes	49
I Words	
2 Regular Expressions and Automata	51
2.1 Regular Expressions	51
2.1.1 Basic Regular Expression Patterns	52
2.1.2 Disjunction, Grouping, and Precedence	55
2.1.3 A Simple Example	56
2.1.4 A More Complex Example	57
2.1.5 Advanced Operators	58
2.1.6 Regular Expression Substitution, Memory, and ELIZA	59
2.2 Finite-State Automata	60
2.2.1 Use of an FSA to Recognize Sheeptalk	61
2.2.2 Formal Languages	64
2.2.3 Another Example	65
2.2.4 Non-Deterministic FSAs	66
2.2.5 Use of an NFSA to Accept Strings	67
2.2.6 Recognition as Search	69
2.2.7 Relation of Deterministic and Non-Deterministic Automata	72
2.3 Regular Languages and FSAs	72
2.4 Summary	75

Bibliographical and Historical Notes	76
Exercises	76
3 Words and Transducers	79
3.1 Survey of (Mostly) English Morphology	81
3.1.1 Inflectional Morphology	82
3.1.2 Derivational Morphology	84
3.1.3 Cliticization	85
3.1.4 Non-Concatenative Morphology	85
3.1.5 Agreement	86
3.2 Finite-State Morphological Parsing	86
3.3 Construction of a Finite-State Lexicon	88
3.4 Finite-State Transducers	91
3.4.1 Sequential Transducers and Determinism	93
3.5 FSTs for Morphological Parsing	94
3.6 Transducers and Orthographic Rules	96
3.7 The Combination of an FST Lexicon and Rules	99
3.8 Lexicon-Free FSTs: The Porter Stemmer	102
3.9 Word and Sentence Tokenization	102
3.9.1 Segmentation in Chinese	104
3.10 Detection and Correction of Spelling Errors	106
3.11 Minimum Edit Distance	107
3.12 Human Morphological Processing	111
3.13 Summary	113
Bibliographical and Historical Notes	114
Exercises	115
4 N-Grams	117
4.1 Word Counting in Corpora	119
4.2 Simple (Unsmoothed) <i>N</i> -Grams	120
4.3 Training and Test Sets	125
4.3.1 <i>N</i> -Gram Sensitivity to the Training Corpus	126
4.3.2 Unknown Words: Open Versus Closed Vocabulary Tasks	129
4.4 Evaluating <i>N</i> -Grams: Perplexity	129
4.5 Smoothing	131
4.5.1 Laplace Smoothing	132
4.5.2 Good-Turing Discounting	135
4.5.3 Some Advanced Issues in Good-Turing Estimation	136
4.6 Interpolation	138
4.7 Backoff	139
4.7.1 Advanced: Details of Computing Katz Backoff α and P^*	141
4.8 Practical Issues: Toolkits and Data Formats	142
4.9 Advanced Issues in Language Modeling	143
4.9.1 Advanced Smoothing Methods: Kneser-Ney Smoothing	143
4.9.2 Class-Based <i>N</i> -Grams	145
4.9.3 Language Model Adaptation and Web Use	146

4.9.4	Using Longer-Distance Information: A Brief Summary . . .	146
4.10	Advanced: Information Theory Background	148
4.10.1	Cross-Entropy for Comparing Models	150
4.11	Advanced: The Entropy of English and Entropy Rate Constancy . .	152
4.12	Summary	153
	Bibliographical and Historical Notes	154
	Exercises	155
5	Part-of-Speech Tagging	157
5.1	(Mostly) English Word Classes	158
5.2	Tagsets for English	164
5.3	Part-of-Speech Tagging	167
5.4	Rule-Based Part-of-Speech Tagging	169
5.5	HMM Part-of-Speech Tagging	173
5.5.1	Computing the Most Likely Tag Sequence: An Example . .	176
5.5.2	Formalizing Hidden Markov Model Taggers	178
5.5.3	Using the Viterbi Algorithm for HMM Tagging	179
5.5.4	Extending the HMM Algorithm to Trigrams	183
5.6	Transformation-Based Tagging	185
5.6.1	How TBL Rules Are Applied	186
5.6.2	How TBL Rules Are Learned	186
5.7	Evaluation and Error Analysis	187
5.7.1	Error Analysis	190
5.8	Advanced Issues in Part-of-Speech Tagging	191
5.8.1	Practical Issues: Tag Indeterminacy and Tokenization	191
5.8.2	Unknown Words	192
5.8.3	Part-of-Speech Tagging for Other Languages	194
5.8.4	Tagger Combination	197
5.9	Advanced: The Noisy Channel Model for Spelling	197
5.9.1	Contextual Spelling Error Correction	201
5.10	Summary	202
	Bibliographical and Historical Notes	203
	Exercises	205
6	Hidden Markov and Maximum Entropy Models	207
6.1	Markov Chains	208
6.2	The Hidden Markov Model	210
6.3	Likelihood Computation: The Forward Algorithm	213
6.4	Decoding: The Viterbi Algorithm	218
6.5	HMM Training: The Forward-Backward Algorithm	220
6.6	Maximum Entropy Models: Background	227
6.6.1	Linear Regression	228
6.6.2	Logistic Regression	231
6.6.3	Logistic Regression: Classification	233
6.6.4	Advanced: Learning in Logistic Regression	234
6.7	Maximum Entropy Modeling	235

6.7.1	Why We Call It Maximum Entropy	239
6.8	Maximum Entropy Markov Models	241
6.8.1	Decoding and Learning in MEMMs	244
6.9	Summary	245
	Bibliographical and Historical Notes	246
	Exercises	247

II Speech

7	Phonetics	249
7.1	Speech Sounds and Phonetic Transcription	250
7.2	Articulatory Phonetics	251
7.2.1	The Vocal Organs	252
7.2.2	Consonants: Place of Articulation	254
7.2.3	Consonants: Manner of Articulation	255
7.2.4	Vowels	256
7.2.5	Syllables	257
7.3	Phonological Categories and Pronunciation Variation	259
7.3.1	Phonetic Features	261
7.3.2	Predicting Phonetic Variation	262
7.3.3	Factors Influencing Phonetic Variation	263
7.4	Acoustic Phonetics and Signals	264
7.4.1	Waves	264
7.4.2	Speech Sound Waves	265
7.4.3	Frequency and Amplitude; Pitch and Loudness	267
7.4.4	Interpretation of Phones from a Waveform	270
7.4.5	Spectra and the Frequency Domain	270
7.4.6	The Source-Filter Model	274
7.5	Phonetic Resources	275
7.6	Advanced: Articulatory and Gestural Phonology	278
7.7	Summary	279
	Bibliographical and Historical Notes	280
	Exercises	281
8	Speech Synthesis	283
8.1	Text Normalization	285
8.1.1	Sentence Tokenization	285
8.1.2	Non-Standard Words	286
8.1.3	Homograph Disambiguation	290
8.2	Phonetic Analysis	291
8.2.1	Dictionary Lookup	291
8.2.2	Names	292
8.2.3	Grapheme-to-Phoneme Conversion	293
8.3	Prosodic Analysis	296
8.3.1	Prosodic Structure	296
8.3.2	Prosodic Prominence	297

8.3.3	Tune	299
8.3.4	More Sophisticated Models: ToBI	300
8.3.5	Computing Duration from Prosodic Labels	302
8.3.6	Computing F0 from Prosodic Labels	303
8.3.7	Final Result of Text Analysis: Internal Representation	305
8.4	Diphone Waveform Synthesis	306
8.4.1	Steps for Building a Diphone Database	306
8.4.2	Diphone Concatenation and TD-PSOLA for Prosody	308
8.5	Unit Selection (Waveform) Synthesis	310
8.6	Evaluation	314
	Bibliographical and Historical Notes	315
	Exercises	318
9	Automatic Speech Recognition	319
9.1	Speech Recognition Architecture	321
9.2	The Hidden Markov Model Applied to Speech	325
9.3	Feature Extraction: MFCC Vectors	329
9.3.1	Preemphasis	330
9.3.2	Windowing	330
9.3.3	Discrete Fourier Transform	332
9.3.4	Mel Filter Bank and Log	333
9.3.5	The Cepstrum: Inverse Discrete Fourier Transform	334
9.3.6	Deltas and Energy	336
9.3.7	Summary: MFCC	336
9.4	Acoustic Likelihood Computation	337
9.4.1	Vector Quantization	337
9.4.2	Gaussian PDFs	340
9.4.3	Probabilities, Log-Probabilities, and Distance Functions	347
9.5	The Lexicon and Language Model	348
9.6	Search and Decoding	348
9.7	Embedded Training	358
9.8	Evaluation: Word Error Rate	362
9.9	Summary	364
	Bibliographical and Historical Notes	365
	Exercises	367
10	Speech Recognition: Advanced Topics	369
10.1	Multipass Decoding: N -Best Lists and Lattices	369
10.2	A^* (“Stack”) Decoding	375
10.3	Context-Dependent Acoustic Models: Triphones	379
10.4	Discriminative Training	383
10.4.1	Maximum Mutual Information Estimation	384
10.4.2	Acoustic Models Based on Posterior Classifiers	385
10.5	Modeling Variation	386
10.5.1	Environmental Variation and Noise	386
10.5.2	Speaker Variation and Speaker Adaptation	387

10.5.3	Pronunciation Modeling: Variation Due to Genre	388
10.6	Metadata: Boundaries, Punctuation, and Disfluencies	390
10.7	Speech Recognition by Humans	392
10.8	Summary	393
	Bibliographical and Historical Notes	393
	Exercises	394
11	Computational Phonology	395
11.1	Finite-State Phonology	395
11.2	Advanced Finite-State Phonology	399
11.2.1	Harmony	399
11.2.2	Templatic Morphology	400
11.3	Computational Optimality Theory	401
11.3.1	Finite-State Transducer Models of Optimality Theory	403
11.3.2	Stochastic Models of Optimality Theory	404
11.4	Syllabification	406
11.5	Learning Phonology and Morphology	409
11.5.1	Learning Phonological Rules	409
11.5.2	Learning Morphology	411
11.5.3	Learning in Optimality Theory	414
11.6	Summary	415
	Bibliographical and Historical Notes	415
	Exercises	417

III Syntax

12	Formal Grammars of English	419
12.1	Constituency	420
12.2	Context-Free Grammars	421
12.2.1	Formal Definition of Context-Free Grammar	425
12.3	Some Grammar Rules for English	426
12.3.1	Sentence-Level Constructions	426
12.3.2	Clauses and Sentences	428
12.3.3	The Noun Phrase	428
12.3.4	Agreement	432
12.3.5	The Verb Phrase and Subcategorization	434
12.3.6	Auxiliaries	436
12.3.7	Coordination	437
12.4	Treebanks	438
12.4.1	Example: The Penn Treebank Project	438
12.4.2	Treebanks as Grammars	440
12.4.3	Treebank Searching	442
12.4.4	Heads and Head Finding	443
12.5	Grammar Equivalence and Normal Form	446
12.6	Finite-State and Context-Free Grammars	447
12.7	Dependency Grammars	448

12.7.1	The Relationship Between Dependencies and Heads	449
12.7.2	Categorial Grammar	451
12.8	Spoken Language Syntax	451
12.8.1	Disfluencies and Repair	452
12.8.2	Treebanks for Spoken Language	453
12.9	Grammars and Human Processing	454
12.10	Summary	455
	Bibliographical and Historical Notes	456
	Exercises	458
13	Syntactic Parsing	461
13.1	Parsing as Search	462
13.1.1	Top-Down Parsing	463
13.1.2	Bottom-Up Parsing	464
13.1.3	Comparing Top-Down and Bottom-Up Parsing	465
13.2	Ambiguity	466
13.3	Search in the Face of Ambiguity	468
13.4	Dynamic Programming Parsing Methods	469
13.4.1	CKY Parsing	470
13.4.2	The Earley Algorithm	477
13.4.3	Chart Parsing	482
13.5	Partial Parsing	484
13.5.1	Finite-State Rule-Based Chunking	486
13.5.2	Machine Learning-Based Approaches to Chunking	486
13.5.3	Chunking-System Evaluations	489
13.6	Summary	490
	Bibliographical and Historical Notes	491
	Exercises	492
14	Statistical Parsing	493
14.1	Probabilistic Context-Free Grammars	494
14.1.1	PCFGs for Disambiguation	495
14.1.2	PCFGs for Language Modeling	497
14.2	Probabilistic CKY Parsing of PCFGs	498
14.3	Ways to Learn PCFG Rule Probabilities	501
14.4	Problems with PCFGs	502
14.4.1	Independence Assumptions Miss Structural Dependencies Be- tween Rules	502
14.4.2	Lack of Sensitivity to Lexical Dependencies	503
14.5	Improving PCFGs by Splitting Non-Terminals	505
14.6	Probabilistic Lexicalized CFGs	507
14.6.1	The Collins Parser	509
14.6.2	Advanced: Further Details of the Collins Parser	511
14.7	Evaluating Parsers	513
14.8	Advanced: Discriminative Reranking	515
14.9	Advanced: Parser-Based Language Modeling	516

14.10	Human Parsing	517
14.11	Summary	519
	Bibliographical and Historical Notes	520
	Exercises	522
15	Features and Unification	523
15.1	Feature Structures	524
15.2	Unification of Feature Structures	526
15.3	Feature Structures in the Grammar	531
15.3.1	Agreement	532
15.3.2	Head Features	534
15.3.3	Subcategorization	535
15.3.4	Long-Distance Dependencies	540
15.4	Implementation of Unification	541
15.4.1	Unification Data Structures	541
15.4.2	The Unification Algorithm	543
15.5	Parsing with Unification Constraints	547
15.5.1	Integration of Unification into an Earley Parser	548
15.5.2	Unification-Based Parsing	553
15.6	Types and Inheritance	555
15.6.1	Advanced: Extensions to Typing	558
15.6.2	Other Extensions to Unification	559
15.7	Summary	559
	Bibliographical and Historical Notes	560
	Exercises	561
16	Language and Complexity	563
16.1	The Chomsky Hierarchy	564
16.2	Ways to Tell if a Language Isn't Regular	566
16.2.1	The Pumping Lemma	567
16.2.2	Proofs that Various Natural Languages Are Not Regular	569
16.3	Is Natural Language Context Free?	571
16.4	Complexity and Human Processing	573
16.5	Summary	576
	Bibliographical and Historical Notes	577
	Exercises	578
 IV Semantics and Pragmatics		
17	The Representation of Meaning	579
17.1	Computational Desiderata for Representations	581
17.1.1	Verifiability	581
17.1.2	Unambiguous Representations	582
17.1.3	Canonical Form	583
17.1.4	Inference and Variables	584
17.1.5	Expressiveness	585

17.2	Model-Theoretic Semantics	586
17.3	First-Order Logic	589
17.3.1	Basic Elements of First-Order Logic	589
17.3.2	Variables and Quantifiers	591
17.3.3	Lambda Notation	593
17.3.4	The Semantics of First-Order Logic	594
17.3.5	Inference	595
17.4	Event and State Representations	597
17.4.1	Representing Time	600
17.4.2	Aspect	603
17.5	Description Logics	606
17.6	Embodied and Situated Approaches to Meaning	612
17.7	Summary	614
	Bibliographical and Historical Notes	614
	Exercises	616
18	Computational Semantics	617
18.1	Syntax-Driven Semantic Analysis	617
18.2	Semantic Augmentations to Syntactic Rules	619
18.3	Quantifier Scope Ambiguity and Underspecification	626
18.3.1	Store and Retrieve Approaches	626
18.3.2	Constraint-Based Approaches	629
18.4	Unification-Based Approaches to Semantic Analysis	632
18.5	Integration of Semantics into the Earley Parser	638
18.6	Idioms and Compositionality	639
18.7	Summary	641
	Bibliographical and Historical Notes	641
	Exercises	643
19	Lexical Semantics	645
19.1	Word Senses	646
19.2	Relations Between Senses	649
19.2.1	Synonymy and Antonymy	649
19.2.2	Hyponymy	650
19.2.3	Semantic Fields	651
19.3	WordNet: A Database of Lexical Relations	651
19.4	Event Participants	653
19.4.1	Thematic Roles	654
19.4.2	Diathesis Alternations	656
19.4.3	Problems with Thematic Roles	657
19.4.4	The Proposition Bank	658
19.4.5	FrameNet	659
19.4.6	Selectional Restrictions	661
19.5	Primitive Decomposition	663
19.6	Advanced: Metaphor	665
19.7	Summary	666

Bibliographical and Historical Notes	667
Exercises	668
20 Computational Lexical Semantics	671
20.1 Word Sense Disambiguation: Overview	672
20.2 Supervised Word Sense Disambiguation	673
20.2.1 Feature Extraction for Supervised Learning	674
20.2.2 Naive Bayes and Decision List Classifiers	675
20.3 WSD Evaluation, Baselines, and Ceilings	678
20.4 WSD: Dictionary and Thesaurus Methods	680
20.4.1 The Lesk Algorithm	680
20.4.2 Selectional Restrictions and Selectional Preferences	682
20.5 Minimally Supervised WSD: Bootstrapping	684
20.6 Word Similarity: Thesaurus Methods	686
20.7 Word Similarity: Distributional Methods	692
20.7.1 Defining a Word's Co-Occurrence Vectors	693
20.7.2 Measuring Association with Context	695
20.7.3 Defining Similarity Between Two Vectors	697
20.7.4 Evaluating Distributional Word Similarity	701
20.8 Hyponymy and Other Word Relations	701
20.9 Semantic Role Labeling	704
20.10 Advanced: Unsupervised Sense Disambiguation	708
20.11 Summary	709
Bibliographical and Historical Notes	710
Exercises	713
21 Computational Discourse	715
21.1 Discourse Segmentation	718
21.1.1 Unsupervised Discourse Segmentation	718
21.1.2 Supervised Discourse Segmentation	720
21.1.3 Discourse Segmentation Evaluation	722
21.2 Text Coherence	723
21.2.1 Rhetorical Structure Theory	724
21.2.2 Automatic Coherence Assignment	726
21.3 Reference Resolution	729
21.4 Reference Phenomena	732
21.4.1 Five Types of Referring Expressions	732
21.4.2 Information Status	734
21.5 Features for Pronominal Anaphora Resolution	735
21.5.1 Features for Filtering Potential Referents	735
21.5.2 Preferences in Pronoun Interpretation	736
21.6 Three Algorithms for Anaphora Resolution	738
21.6.1 Pronominal Anaphora Baseline: The Hobbs Algorithm	738
21.6.2 A Centering Algorithm for Anaphora Resolution	740
21.6.3 A Log-Linear Model for Pronominal Anaphora Resolution	742
21.6.4 Features for Pronominal Anaphora Resolution	743

21.7	Coreference Resolution	744
21.8	Evaluation of Coreference Resolution	746
21.9	Advanced: Inference-Based Coherence Resolution	747
21.10	Psycholinguistic Studies of Reference	752
21.11	Summary	753
	Bibliographical and Historical Notes	754
	Exercises	756

V Applications

22	Information Extraction	759
22.1	Named Entity Recognition	761
22.1.1	Ambiguity in Named Entity Recognition	763
22.1.2	NER as Sequence Labeling	763
22.1.3	Evaluation of Named Entity Recognition	766
22.1.4	Practical NER Architectures	768
22.2	Relation Detection and Classification	768
22.2.1	Supervised Learning Approaches to Relation Analysis	769
22.2.2	Lightly Supervised Approaches to Relation Analysis	772
22.2.3	Evaluation of Relation Analysis Systems	776
22.3	Temporal and Event Processing	777
22.3.1	Temporal Expression Recognition	777
22.3.2	Temporal Normalization	780
22.3.3	Event Detection and Analysis	783
22.3.4	TimeBank	784
22.4	Template Filling	786
22.4.1	Statistical Approaches to Template-Filling	786
22.4.2	Finite-State Template-Filling Systems	788
22.5	Advanced: Biomedical Information Extraction	791
22.5.1	Biological Named Entity Recognition	792
22.5.2	Gene Normalization	793
22.5.3	Biological Roles and Relations	794
22.6	Summary	796
	Bibliographical and Historical Notes	796
	Exercises	797
23	Question Answering and Summarization	799
23.1	Information Retrieval	801
23.1.1	The Vector Space Model	802
23.1.2	Term Weighting	804
23.1.3	Term Selection and Creation	806
23.1.4	Evaluation of Information-Retrieval Systems	806
23.1.5	Homonymy, Polysemy, and Synonymy	810
23.1.6	Ways to Improve User Queries	810
23.2	Factoid Question Answering	812
23.2.1	Question Processing	813

23.2.2	Passage Retrieval	815
23.2.3	Answer Processing	817
23.2.4	Evaluation of Factoid Answers	821
23.3	Summarization	821
23.4	Single-Document Summarization	824
23.4.1	Unsupervised Content Selection	824
23.4.2	Unsupervised Summarization Based on Rhetorical Parsing	826
23.4.3	Supervised Content Selection	828
23.4.4	Sentence Simplification	829
23.5	Multi-Document Summarization	830
23.5.1	Content Selection in Multi-Document Summarization	831
23.5.2	Information Ordering in Multi-Document Summarization	832
23.6	Focused Summarization and Question Answering	835
23.7	Summarization Evaluation	839
23.8	Summary	841
	Bibliographical and Historical Notes	842
	Exercises	844

24 Dialogue and Conversational Agents 847

24.1	Properties of Human Conversations	849
24.1.1	Turns and Turn-Taking	849
24.1.2	Language as Action: Speech Acts	851
24.1.3	Language as Joint Action: Grounding	852
24.1.4	Conversational Structure	854
24.1.5	Conversational Implicature	855
24.2	Basic Dialogue Systems	857
24.2.1	ASR Component	857
24.2.2	NLU Component	858
24.2.3	Generation and TTS Components	861
24.2.4	Dialogue Manager	863
24.2.5	Dealing with Errors: Confirmation and Rejection	867
24.3	VoiceXML	868
24.4	Dialogue System Design and Evaluation	872
24.4.1	Designing Dialogue Systems	872
24.4.2	Evaluating Dialogue Systems	872
24.5	Information-State and Dialogue Acts	874
24.5.1	Using Dialogue Acts	876
24.5.2	Interpreting Dialogue Acts	877
24.5.3	Detecting Correction Acts	880
24.5.4	Generating Dialogue Acts: Confirmation and Rejection	881
24.6	Markov Decision Process Architecture	882
24.7	Advanced: Plan-Based Dialogue Agents	886
24.7.1	Plan-Inferential Interpretation and Production	887
24.7.2	The Intentional Structure of Dialogue	889
24.8	Summary	891
	Bibliographical and Historical Notes	892

Exercises	894
25 Machine Translation	895
25.1 Why Machine Translation Is Hard	898
25.1.1 Typology	898
25.1.2 Other Structural Divergences	900
25.1.3 Lexical Divergences	901
25.2 Classical MT and the Vauquois Triangle	903
25.2.1 Direct Translation	904
25.2.2 Transfer	906
25.2.3 Combined Direct and Transfer Approaches in Classic MT	908
25.2.4 The Interlingua Idea: Using Meaning	909
25.3 Statistical MT	910
25.4 $P(F E)$: The Phrase-Based Translation Model	913
25.5 Alignment in MT	915
25.5.1 IBM Model 1	916
25.5.2 HMM Alignment	919
25.6 Training Alignment Models	921
25.6.1 EM for Training Alignment Models	922
25.7 Symmetrizing Alignments for Phrase-Based MT	924
25.8 Decoding for Phrase-Based Statistical MT	926
25.9 MT Evaluation	930
25.9.1 Using Human Raters	930
25.9.2 Automatic Evaluation: BLEU	931
25.10 Advanced: Syntactic Models for MT	934
25.11 Advanced: IBM Model 3 and Fertility	935
25.11.1 Training for Model 3	939
25.12 Advanced: Log-Linear Models for MT	939
25.13 Summary	940
Bibliographical and Historical Notes	941
Exercises	943
Bibliography	945
Author Index	995
Subject Index	1007