

# **SPOKEN LANGUAGE UNDERSTANDING**

## **SYSTEMS FOR EXTRACTING SEMANTIC INFORMATION FROM SPEECH**

**Gokhan Tur**

*Microsoft Speech Labs, Microsoft Research, USA*

**Renato De Mori**

*McGill University, Montreal, Canada and University of Avignon, France*



A John Wiley and Sons, Ltd, Publication

# Contents

<b>List of Contributors</b>	<b>xvii</b>
<b>Foreword</b>	<b>xxv</b>
<b>Preface</b>	<b>xxix</b>
<b>1 Introduction</b>	<b>1</b>
<i>Gokhan Tur and Renato De Mori</i>	
1.1 A Brief History of Spoken Language Understanding	1
1.2 Organization of the Book	4
1.2.1 <i>Part I. Spoken Language Understanding for Human/Machine Interactions</i>	4
1.2.2 <i>Part II. Spoken Language Understanding for Human/Human Conversations</i>	6
References	7
 <b>PART 1 SPOKEN LANGUAGE UNDERSTANDING FOR HUMAN/MACHINE INTERACTIONS</b>	
<b>2 History of Knowledge and Processes for Spoken Language Understanding</b>	<b>11</b>
<i>Renato De Mori</i>	
2.1 Introduction	11
2.2 Meaning Representation and Sentence Interpretation	12
2.2.1 <i>Meaning Representation Languages</i>	12
2.2.2 <i>Meaning Extraction from Sentences</i>	16
2.3 Knowledge Fragments and Semantic Composition	18
2.3.1 <i>Concept Tags and Knowledge Fragments</i>	19
2.3.2 <i>Composition by Fusion of Fragments</i>	21
2.3.3 <i>Composition by Attachment</i>	23
2.3.4 <i>Composition by Attachment and Inference</i>	24
2.4 Probabilistic Interpretation in SLU Systems	25
2.5 Interpretation with Partial Syntactic Analysis	26
2.6 Classification Models for Interpretation	28
2.7 Advanced Methods and Resources for Semantic Modeling and Interpretation	30
2.8 Recent Systems	32
2.9 Conclusions	35
References	36

<b>3</b>	<b>Semantic Frame-based Spoken Language Understanding</b>	<b>41</b>
	<i>Ye-Yi Wang, Li Deng and Alex Acero</i>	
3.1	Background	41
3.1.1	<i>History of the Frame-based SLU</i>	41
3.1.2	<i>Semantic Representation and Semantic Frame</i>	43
3.1.3	<i>Technical Challenges</i>	45
3.1.4	<i>Standard Data Sets</i>	47
3.1.5	<i>Evaluation Metrics</i>	47
3.2	Knowledge-based Solutions	49
3.2.1	<i>Semantically Enhanced Syntactic Grammars</i>	49
3.2.2	<i>Semantic Grammars</i>	51
3.2.3	<i>Knowledge-based Solutions in Commercial Applications</i>	52
3.3	Data-driven Approaches	54
3.3.1	<i>Generative Models</i>	55
3.3.2	<i>Integrating Knowledge in Statistical Models – A Case Study of the Generative HMM/CFG Composite Model</i>	65
3.3.3	<i>Use of Generative Understanding Models in Speech Recognition</i>	71
3.3.4	<i>Conditional Models</i>	74
3.3.5	<i>Other Data-driven Approaches to SLU</i>	84
3.3.6	<i>Frame-based SLU in Context</i>	86
3.4	Summary	87
	References	88
<b>4</b>	<b>Intent Determination and Spoken Utterance Classification</b>	<b>93</b>
	<i>Gokhan Tur and Li Deng</i>	
4.1	Background	93
4.2	Task Description	96
4.3	Technical Challenges	97
4.4	Benchmark Data Sets	98
4.5	Evaluation Metrics	98
4.5.1	<i>Direct Metrics</i>	98
4.5.2	<i>Indirect Metrics</i>	99
4.6	Technical Approaches	99
4.6.1	<i>Semantic Representations</i>	100
4.6.2	<i>The HMIHY Way: Using Salient Phrases</i>	101
4.6.3	<i>Vector-state Model</i>	103
4.6.4	<i>Using Discriminative Classifiers</i>	103
4.6.5	<i>Using Prior Knowledge</i>	105
4.6.6	<i>Beyond ASR 1-Best: Using Word Confusion Networks</i>	106
4.6.7	<i>Conditional Understanding Models Used for Discriminative Training of Language Models</i>	108
4.6.8	<i>Phone-based Call Classification</i>	115
4.7	Discussion and Conclusions	115
	References	117

<b>5</b>	<b>Voice Search</b>	<b>119</b>
	<i>Ye-Yi Wang, Dong Yu, Yun-Cheng Ju and Alex Acero</i>	
5.1	Background	119
5.1.1	<i>Voice Search Compared with the Other Spoken Dialogue Technologies</i>	120
5.1.2	<i>History of Voice Search</i>	122
5.1.3	<i>Technical Challenges</i>	124
5.1.4	<i>Data Sets</i>	125
5.1.5	<i>Evaluation Metrics</i>	125
5.2	Technology Review	128
5.2.1	<i>Speech Recognition</i>	128
5.2.2	<i>Spoken Language Understanding/Search</i>	133
5.2.3	<i>Dialogue Management</i>	140
5.2.4	<i>Closing the Feedback Loop</i>	143
5.3	Summary	144
	References	144
<b>6</b>	<b>Spoken Question Answering</b>	<b>147</b>
	<i>Sophie Rosset, Olivier Galibert and Lori Lamel</i>	
6.1	Introduction	147
6.2	Specific Aspects of Handling Speech in QA Systems	149
6.3	QA Evaluation Campaigns	150
6.3.1	<i>General Presentation</i>	151
6.3.2	<i>Question Answering on Speech Transcripts: Evaluation Campaigns</i>	154
6.4	Question-answering Systems	156
6.4.1	<i>General Overview</i>	156
6.4.2	<i>Approaches Used in the QAsT Campaigns</i>	158
6.4.3	<i>QAsT Campaign Results</i>	162
6.5	Projects Integrating Spoken Requests and Question Answering	166
6.6	Conclusions	167
	References	167
<b>7</b>	<b>SLU in Commercial and Research Spoken Dialogue Systems</b>	<b>171</b>
	<i>David Suendermann and Roberto Pieraccini</i>	
7.1	Why Spoken Dialogue Systems do not have to Understand	171
7.2	Approaches to SLU for Dialogue Systems	173
7.2.1	<i>Rule-based Semantic Grammars</i>	174
7.2.2	<i>Statistical SLU</i>	175
7.2.3	<i>Dealing with Deficiencies of Speech Recognition and SLU in Dialogue Systems</i>	177
7.2.4	<i>Robust Interaction Design and Multiple Levels of Confidence Thresholds</i>	177
7.2.5	<i>N-best Lists</i>	178
7.2.6	<i>One-step Correction and Mixed Initiative</i>	179
7.2.7	<i>Belief Systems</i>	180

7.3	From Call Flow to POMDP: How Dialogue Management Integrates with SLU	180
7.3.1	<i>Rule-based Approaches: Call Flow, Form-filling, Agenda, Call-routing, Inference</i>	181
7.3.2	<i>Statistical Dialogue Management: Reinforcement Learning, MDP, POMDP</i>	183
7.4	Benchmark Projects and Data Sets	186
7.4.1	<i>ATIS</i>	186
7.4.2	<i>Communicator</i>	186
7.4.3	<i>Let's Go!</i>	187
7.4.4	<i>Datasets in Commercial Dialogue Systems</i>	187
7.5	Time is Money: The Relationship between SLU and Overall Dialogue System Performance	189
7.5.1	<i>Automation Rate</i>	189
7.5.2	<i>Average Handling Time</i>	190
7.5.3	<i>Retry Rate and Speech Errors</i>	190
7.6	Conclusion	191
	References	191
<b>8</b>	<b>Active Learning</b>	<b>195</b>
	<i>Dilek Hakkani-Tür and Giuseppe Riccardi</i>	
8.1	Introduction	195
8.2	Motivation	196
8.2.1	<i>Language Variability</i>	196
8.2.2	<i>The Domain Concept Variability</i>	198
8.2.3	<i>Noisy Annotation</i>	200
8.2.4	<i>The Data Overflow</i>	201
8.3	Learning Architectures	201
8.3.1	<i>Passive Learning</i>	201
8.3.2	<i>Active Learning</i>	202
8.4	Active Learning Methods	204
8.4.1	<i>The Statistical Framework</i>	204
8.4.2	<i>Certainty-based Active Learning Methods</i>	205
8.4.3	<i>Committee-based Active Learning</i>	208
8.4.4	<i>Density-based Active Learning</i>	209
8.4.5	<i>Stopping Criteria for Active Learning</i>	211
8.5	Combining Active Learning with Semi-supervised Learning	211
8.6	Applications	213
8.6.1	<i>Automatic Speech Recognition</i>	213
8.6.2	<i>Intent Determination</i>	215
8.6.3	<i>Concept Segmentation/Labeling</i>	217
8.6.4	<i>Dialogue Act Tagging</i>	218
8.7	Evaluation of Active Learning Methods	219
8.8	Discussion and Conclusions	220
	References	221

## PART 2 SPOKEN LANGUAGE UNDERSTANDING FOR HUMAN/HUMAN CONVERSATIONS

<b>9 Human/Human Conversation Understanding</b>	<b>227</b>
<i>Gokhan Tur and Dilek Hakkani-Tür</i>	
9.1 Background	227
9.2 Human/Human Conversation Understanding Tasks	229
9.3 Dialogue Act Segmentation and Tagging	231
9.3.1 Annotation Schema	232
9.3.2 Modeling Dialogue Act Tagging	236
9.3.3 Dialogue Act Segmentation	237
9.3.4 Joint Modeling of Dialogue Act Segmentation and Tagging	239
9.4 Action Item and Decision Detection	240
9.5 Addressee Detection and Co-reference Resolution	242
9.6 Hot Spot Detection	244
9.7 Subjectivity, Sentiment, and Opinion Detection	244
9.8 Speaker Role Detection	245
9.9 Modeling Dominance	247
9.10 Argument Diagramming	247
9.11 Discussion and Conclusions	250
References	251
 <b>10 Named Entity Recognition</b>	 <b>257</b>
<i>Frédéric Béchet</i>	
10.1 Task Description	258
10.1.1 What is a Named Entity?	258
10.1.2 What are the Main Issues in the NER Task?	260
10.1.3 Applicative Frameworks of NER in Speech	261
10.2 Challenges Using Speech Input	263
10.3 Benchmark Data Sets, Applications	265
10.3.1 NER as an IE Task	265
10.3.2 NER as an SLU Task in a Spoken Dialogue Context	266
10.4 Evaluation Metrics	266
10.4.1 Aligning the Reference and Hypothesis NE Annotations	267
10.4.2 Scoring	267
10.5 Main Approaches for Extracting NEs from Text	269
10.5.1 Rules and Grammars	269
10.5.2 NER as a Word Tagging Problem	270
10.5.3 Hidden Markov Model	271
10.5.4 Maximum Entropy	273
10.5.5 Conditional Random Field	274
10.5.6 Sample Classification Methods	275
10.5.7 Conclusions on the Methods for NER from Text	276
10.6 Comparative Methods for NER from Speech	277
10.6.1 Adapting NER Systems to ASR Output	277
10.6.2 Integrating ASR and NER Processes	281

10.7	New Trends in NER from Speech	284
10.7.1	<i>Adapting the ASR Lexicon</i>	284
10.7.2	<i>Collecting Data on the ASR Lexicon</i>	285
10.7.3	<i>Toward an Open-vocabulary ASR System for NER from Speech</i>	286
10.8	Conclusions	287
	References	287
<b>11</b>	<b>Topic Segmentation</b>	<b>291</b>
	<i>Matthew Purver</i>	
11.1	Task Description	291
11.1.1	<i>Introduction</i>	291
11.1.2	<i>What is a Topic?</i>	292
11.1.3	<i>Linear versus Hierarchical Segmentation</i>	292
11.2	Basic Approaches, and the Challenge of Speech	293
11.2.1	<i>Changes in Content</i>	293
11.2.2	<i>Distinctive Boundary Features</i>	294
11.2.3	<i>Monologue</i>	294
11.2.4	<i>Dialogue</i>	295
11.3	Applications and Benchmark Datasets	295
11.3.1	<i>Monologue</i>	296
11.3.2	<i>Dialogue</i>	296
11.4	Evaluation Metrics	297
11.4.1	<i>Classification-based</i>	297
11.4.2	<i>Segmentation-based</i>	298
11.4.3	<i>Content-based</i>	302
11.5	Technical Approaches	302
11.5.1	<i>Changes in Lexical Similarity</i>	302
11.5.2	<i>Similarity-based Clustering</i>	305
11.5.3	<i>Generative Models</i>	306
11.5.4	<i>Discriminative Boundary Detection</i>	310
11.5.5	<i>Combined Approaches, and the State of the Art</i>	310
11.6	New Trends and Future Directions	313
11.6.1	<i>Multi-modality</i>	313
11.6.2	<i>Topic Identification and Adaptation</i>	313
	References	314
<b>12</b>	<b>Topic Identification</b>	<b>319</b>
	<i>Timothy J. Hazen</i>	
12.1	Task Description	319
12.1.1	<i>What is Topic Identification?</i>	319
12.1.2	<i>What are Topics?</i>	320
12.1.3	<i>How is Topic Relevancy Defined?</i>	321
12.1.4	<i>Characterizing the Constraints on Topic ID Tasks</i>	321
12.1.5	<i>Text-based Topic Identification</i>	323

12.2	Challenges Using Speech Input	323
12.2.1	<i>The Naive Approach to Speech-based Topic ID</i>	323
12.2.2	<i>Challenges of Extemporaneous Speech</i>	323
12.2.3	<i>Challenges of Imperfect Speech Recognition</i>	324
12.2.4	<i>Challenges of Unconstrained Domains</i>	325
12.3	Applications and Benchmark Tasks	326
12.3.1	<i>The TDT Project</i>	326
12.3.2	<i>The Switchboard and Fisher Corpora</i>	327
12.3.3	<i>Customer Service/Call Routing Applications</i>	327
12.4	Evaluation Metrics	328
12.4.1	<i>Topic Scoring</i>	328
12.4.2	<i>Classification Error Rate</i>	328
12.4.3	<i>Detection-based Evaluation Metrics</i>	328
12.5	Technical Approaches	333
12.5.1	<i>Topic ID System Overview</i>	333
12.5.2	<i>Automatic Speech Recognition</i>	333
12.5.3	<i>Feature Extraction</i>	334
12.5.4	<i>Feature Selection and Transformation</i>	335
12.5.5	<i>Latent Concept Modeling</i>	340
12.5.6	<i>Topic ID Classification and Detection</i>	343
12.5.7	<i>Example Topic ID Results on the Fisher Corpus</i>	346
12.5.8	<i>Novel Topic Detection</i>	350
12.5.9	<i>Topic Clustering</i>	350
12.6	New Trends and Future Directions	352
	References	353
<b>13</b>	<b>Speech Summarization</b>	<b>357</b>
	<i>Yang Liu and Dilek Hakkani-Tür</i>	
13.1	Task Description	357
13.1.1	<i>General Definition of Summarization</i>	357
13.1.2	<i>Speech Summarization</i>	359
13.1.3	<i>Applications</i>	361
13.2	Challenges when Using Speech Input	362
13.2.1	<i>Automatic Speech Recognition Errors</i>	363
13.2.2	<i>Speaker Turns</i>	363
13.2.3	<i>Sentence Boundaries</i>	363
13.2.4	<i>Disfluencies and Ungrammatical Utterances</i>	364
13.2.5	<i>Other Style and Structural Information</i>	365
13.3	Data Sets	366
13.3.1	<i>Broadcast News (BN)</i>	367
13.3.2	<i>Lectures</i>	368
13.3.3	<i>Multi-party Conversational Speech</i>	369
13.3.4	<i>Voice Mail</i>	371
13.4	Evaluation Metrics	371
13.4.1	<i>Recall, Precision, and F-measure</i>	372
13.4.2	<i>ROUGE</i>	372



13.4.3	<i>The Pyramid Method</i>	373
13.4.4	<i>Weighted Precision</i>	374
13.4.5	<i>SumACCY and Weighted SumACCY</i>	374
13.4.6	<i>Human Evaluation</i>	375
13.4.7	<i>Issues and Discussions</i>	375
13.5	<b>General Approaches</b>	375
13.5.1	<i>Extractive Summarization: Unsupervised Methods</i>	376
13.5.2	<i>Extractive Summarization: Supervised Learning Methods</i>	381
13.5.3	<i>Moving Beyond Generic Extractive Summarization</i>	385
13.5.4	<i>Summary</i>	386
13.6	<b>More Discussions on Speech versus Text Summarization</b>	386
13.6.1	<i>Speech Recognition Errors</i>	386
13.6.2	<i>Sentence Segmentation</i>	388
13.6.3	<i>Disfluencies</i>	389
13.6.4	<i>Acoustic/Prosodic and Other Speech Features</i>	390
13.7	<b>Conclusions</b>	391
	<b>References</b>	392
<b>14</b>	<b>Speech Analytics</b>	<b>397</b>
	<i>I. Dan Melamed and Mazin Gilbert</i>	
14.1	<b>Introduction</b>	397
14.2	<b>System Architecture</b>	398
14.3	<b>Speech Transcription</b>	401
14.4	<b>Text Feature Extraction</b>	402
14.5	<b>Acoustic Feature Extraction</b>	403
14.6	<b>Relational Feature Extraction</b>	405
14.7	<b>DBMS</b>	405
14.8	<b>Media Server and Player</b>	408
14.9	<b>Trend Analysis</b>	409
14.10	<b>Alerting System</b>	413
14.11	<b>Conclusion</b>	414
	<b>References</b>	415
<b>15</b>	<b>Speech Retrieval</b>	<b>417</b>
	<i>Ciprian Chelba, Timothy J. Hazen, Bhuvana Ramabhadran and Murat Saraçlar</i>	
15.1	<b>Task Description</b>	417
15.1.1	<i>Spoken Document Retrieval</i>	417
15.1.2	<i>Spoken Utterance Retrieval</i>	418
15.1.3	<i>Spoken Term Detection</i>	418
15.1.4	<i>Browsing</i>	418
15.2	<b>Applications</b>	418
15.2.1	<i>Broadcast News</i>	419
15.2.2	<i>Academic Lectures</i>	419
15.2.3	<i>Sign Language Video</i>	419
15.2.4	<i>Historical Interviews</i>	420
15.2.5	<i>General Web Video</i>	420

---

15.3	Challenges Using Speech Input	420
15.3.1	<i>Overview</i>	420
15.3.2	<i>Coping with ASR Errors Using Lattices</i>	421
15.3.3	<i>Out-of-vocabulary Words</i>	422
15.3.4	<i>Morphologically Rich Languages</i>	423
15.3.5	<i>Resource-limited Languages and Dialects</i>	423
15.4	Evaluation Metrics	424
15.5	Benchmark Data Sets	425
15.5.1	<i>TREC</i>	425
15.5.2	<i>NIST STD</i>	426
15.6	Approaches	426
15.6.1	<i>Basic SDR Approaches</i>	426
15.6.2	<i>Basic STD Approaches</i>	428
15.6.3	<i>Using Sub-word Units</i>	430
15.6.4	<i>Using Lattices</i>	432
15.6.5	<i>Hybrid and Combination Methods</i>	434
15.6.6	<i>Determining Thresholds</i>	435
15.6.7	<i>Presentation and Browsing</i>	437
15.6.8	<i>Other Previous Work</i>	438
15.7	New Trends	439
15.7.1	<i>Indexing and Retrieval for very Large Corpora</i>	439
15.7.2	<i>Query by Example</i>	441
15.7.3	<i>Optimizing Evaluation Performance</i>	442
15.7.4	<i>Multilingual Speech Retrieval</i>	443
15.8	Discussion and Conclusions	443
	References	444
	<b>Index</b>	<b>447</b>