# Contents

List of Contributors xvii  
Foreword xxv  
Preface xxix  

1 Introduction 1  
   Gokhan Tur and Renato De Mori  
   1.1 A Brief History of Spoken Language Understanding 1  
   1.2 Organization of the Book 4  
      1.2.1 Part I. Spoken Language Understanding for Human/Machine Interactions 4  
      1.2.2 Part II. Spoken Language Understanding for Human/Human Conversations 6  
   References 7  

PART 1  SPOKEN LANGUAGE UNDERSTANDING FOR HUMAN/MACHINE INTERACTIONS  

2 History of Knowledge and Processes for Spoken Language Understanding 11  
   Renato De Mori  
   2.1 Introduction 11  
   2.2 Meaning Representation and Sentence Interpretation 12  
      2.2.1 Meaning Representation Languages 12  
      2.2.2 Meaning Extraction from Sentences 16  
   2.3 Knowledge Fragments and Semantic Composition 18  
      2.3.1 Concept Tags and Knowledge Fragments 19  
      2.3.2 Composition by Fusion of Fragments 21  
      2.3.3 Composition by Attachment 23  
      2.3.4 Composition by Attachment and Inference 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
## 3 Semantic Frame-based Spoken Language Understanding

*Ye-Yi Wang, Li Deng and Alex Acero*

3.1 Background

- 3.1.1 *History of the Frame-based SLU*
- 3.1.2 *Semantic Representation and Semantic Frame*
- 3.1.3 *Technical Challenges*
- 3.1.4 *Standard Data Sets*
- 3.1.5 *Evaluation Metrics*

3.2 Knowledge-based Solutions

- 3.2.1 *Semantically Enhanced Syntactic Grammars*
- 3.2.2 *Semantic Grammars*
- 3.2.3 *Knowledge-based Solutions in Commercial Applications*

3.3 Data-driven Approaches

- 3.3.1 *Generative Models*
- 3.3.2 *Integrating Knowledge in Statistical Models – A Case Study of the Generative HMM/CFG Composite Model*
- 3.3.3 *Use of Generative Understanding Models in Speech Recognition*
- 3.3.4 *Conditional Models*
- 3.3.5 *Other Data-driven Approaches to SLU*
- 3.3.6 *Frame-based SLU in Context*

3.4 Summary

References

## 4 Intent Determination and Spoken Utterance Classification

*Gokhan Tur and Li Deng*

4.1 Background

4.2 Task Description

4.3 Technical Challenges

4.4 Benchmark Data Sets

4.5 Evaluation Metrics

- 4.5.1 *Direct Metrics*
- 4.5.2 *Indirect Metrics*

4.6 Technical Approaches

- 4.6.1 *Semantic Representations*
- 4.6.2 *The HMIHY Way: Using Salient Phrases*
- 4.6.3 *Vector-state Model*
- 4.6.4 *Using Discriminative Classifiers*
- 4.6.5 *Using Prior Knowledge*
- 4.6.6 *Beyond ASR 1-Best: Using Word Confusion Networks*
- 4.6.7 *Conditional Understanding Models Used for Discriminative Training of Language Models*
- 4.6.8 *Phone-based Call Classification*

4.7 Discussion and Conclusions

References
5 Voice Search
Ye-Yi Wang, Dong Yu, Yun-Cheng Ju and Alex Acero
5.1 Background
5.1.1 Voice Search Compared with the Other Spoken Dialogue Technologies
5.1.2 History of Voice Search
5.1.3 Technical Challenges
5.1.4 Data Sets
5.1.5 Evaluation Metrics
5.2 Technology Review
5.2.1 Speech Recognition
5.2.2 Spoken Language Understanding/Search
5.2.3 Dialogue Management
5.2.4 Closing the Feedback Loop
5.3 Summary
References

6 Spoken Question Answering
Sophie Rosset, Olivier Galibert and Lori Lamel
6.1 Introduction
6.2 Specific Aspects of Handling Speech in QA Systems
6.3 QA Evaluation Campaigns
6.3.1 General Presentation
6.3.2 Question Answering on Speech Transcripts: Evaluation Campaigns
6.4 Question-answering Systems
6.4.1 General Overview
6.4.2 Approaches Used in the QAst Campaigns
6.4.3 QAst Campaign Results
6.5 Projects Integrating Spoken Requests and Question Answering
6.6 Conclusions
References

7 SLU in Commercial and Research Spoken Dialogue Systems
David Suendermann and Roberto Pieraccini
7.1 Why Spoken Dialogue Systems do not have to Understand
7.2 Approaches to SLU for Dialogue Systems
7.2.1 Rule-based Semantic Grammars
7.2.2 Statistical SLU
7.2.3 Dealing with Deficiencies of Speech Recognition and SLU in Dialogue Systems
7.2.4 Robust Interaction Design and Multiple Levels of Confidence Thresholds
7.2.5 N-best Lists
7.2.6 One-step Correction and Mixed Initiative
7.2.7 Belief Systems
7.3 From Call Flow to POMDP: How Dialogue Management Integrates with SLU 180
  7.3.1 Rule-based Approaches: Call Flow, Form-filling, Agenda, Call-routing, Inference 181
  7.3.2 Statistical Dialogue Management: Reinforcement Learning, MDP, POMDP 183
7.4 Benchmark Projects and Data Sets 186
  7.4.1 ATIS 186
  7.4.2 Communicator 186
  7.4.3 Let’s Go! 187
  7.4.4 Datasets in Commercial Dialogue Systems 187
7.5 Time is Money: The Relationship between SLU and Overall Dialogue System Performance 189
  7.5.1 Automation Rate 189
  7.5.2 Average Handling Time 190
  7.5.3 Retry Rate and Speech Errors 190
7.6 Conclusion 191
References 191

8 Active Learning 195
Dilek Hakkani-Tür and Giuseppe Riccardi
8.1 Introduction 195
8.2 Motivation 196
  8.2.1 Language Variability 196
  8.2.2 The Domain Concept Variability 198
  8.2.3 Noisy Annotation 200
  8.2.4 The Data Overflow 201
8.3 Learning Architectures 201
  8.3.1 Passive Learning 201
  8.3.2 Active Learning 202
8.4 Active Learning Methods 204
  8.4.1 The Statistical Framework 204
  8.4.2 Certainty-based Active Learning Methods 205
  8.4.3 Committee-based Active Learning 208
  8.4.4 Density-based Active Learning 209
  8.4.5 Stopping Criteria for Active Learning 211
8.5 Combining Active Learning with Semi-supervised Learning 211
8.6 Applications 213
  8.6.1 Automatic Speech Recognition 213
  8.6.2 Intent Determination 215
  8.6.3 Concept Segmentation/Labeling 217
  8.6.4 Dialogue Act Tagging 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

## 9  Human/Human Conversation Understanding  
Gokhan Tur and Dilek Hakkani-Tür  
9.1 Background  
9.2 Human/Human Conversation Understanding Tasks  
9.3 Dialogue Act Segmentation and Tagging  
  9.3.1 Annotation Schema  
  9.3.2 Modeling Dialogue Act Tagging  
  9.3.3 Dialogue Act Segmentation  
  9.3.4 Joint Modeling of Dialogue Act Segmentation and Tagging  
9.4 Action Item and Decision Detection  
9.5 Addressee Detection and Co-reference Resolution  
9.6 Hot Spot Detection  
9.7 Subjectivity, Sentiment, and Opinion Detection  
9.8 Speaker Role Detection  
9.9 Modeling Dominance  
9.10 Argument Diagramming  
9.11 Discussion and Conclusions  
References  

## 10  Named Entity Recognition  
Frédéric Béchet  
10.1 Task Description  
  10.1.1 What is a Named Entity?  
  10.1.2 What are the Main Issues in the NER Task?  
  10.1.3 Applicative Frameworks of NER in Speech  
10.2 Challenges Using Speech Input  
10.3 Benchmark Data Sets, Applications  
  10.3.1 NER as an IE Task  
  10.3.2 NER as an SLU Task in a Spoken Dialogue Context  
10.4 Evaluation Metrics  
  10.4.1 Aligning the Reference and Hypothesis NE Annotations  
  10.4.2 Scoring  
10.5 Main Approaches for Extracting NEs from Text  
  10.5.1 Rules and Grammars  
  10.5.2 NER as a Word Tagging Problem  
  10.5.3 Hidden Markov Model  
  10.5.4 Maximum Entropy  
  10.5.5 Conditional Random Field  
  10.5.6 Sample Classification Methods  
  10.5.7 Conclusions on the Methods for NER from Text  
10.6 Comparative Methods for NER from Speech  
  10.6.1 Adapting NER Systems to ASR Output  
  10.6.2 Integrating ASR and NER Processes  
References
10.7 New Trends in NER from Speech

10.7.1 Adapting the ASR Lexicon

10.7.2 Collecting Data on the ASR Lexicon

10.7.3 Toward an Open-vocabulary ASR System for NER from Speech

10.8 Conclusions

References

11 Topic Segmentation

Matthew Purver

11.1 Task Description

11.1.1 Introduction

11.1.2 What is a Topic?

11.1.3 Linear versus Hierarchical Segmentation

11.2 Basic Approaches, and the Challenge of Speech

11.2.1 Changes in Content

11.2.2 Distinctive Boundary Features

11.2.3 Monologue

11.2.4 Dialogue

11.3 Applications and Benchmark Datasets

11.3.1 Monologue

11.3.2 Dialogue

11.4 Evaluation Metrics

11.4.1 Classification-based

11.4.2 Segmentation-based

11.4.3 Content-based

11.5 Technical Approaches

11.5.1 Changes in Lexical Similarity

11.5.2 Similarity-based Clustering

11.5.3 Generative Models

11.5.4 Discriminative Boundary Detection

11.5.5 Combined Approaches, and the State of the Art

11.6 New Trends and Future Directions

11.6.1 Multi-modality

11.6.2 Topic Identification and Adaptation

References

12 Topic Identification

Timothy J. Hazen

12.1 Task Description

12.1.1 What is Topic Identification?

12.1.2 What are Topics?

12.1.3 How is Topic Relevancy Defined?

12.1.4 Characterizing the Constraints on Topic ID Tasks

12.1.5 Text-based Topic Identification

References
12.2 Challenges Using Speech Input  323
  12.2.1 The Naive Approach to Speech-based Topic ID  323
  12.2.2 Challenges of Extemporaneous Speech  323
  12.2.3 Challenges of Imperfect Speech Recognition  324
  12.2.4 Challenges of Unconstrained Domains  325
12.3 Applications and Benchmark Tasks  326
  12.3.1 The TDT Project  326
  12.3.2 The Switchboard and Fisher Corpora  327
  12.3.3 Customer Service/Call Routing Applications  327
12.4 Evaluation Metrics  328
  12.4.1 Topic Scoring  328
  12.4.2 Classification Error Rate  328
  12.4.3 Detection-based Evaluation Metrics  328
12.5 Technical Approaches  333
  12.5.1 Topic ID System Overview  333
  12.5.2 Automatic Speech Recognition  333
  12.5.3 Feature Extraction  334
  12.5.4 Feature Selection and Transformation  335
  12.5.5 Latent Concept Modeling  340
  12.5.6 Topic ID Classification and Detection  343
  12.5.7 Example Topic ID Results on the Fisher Corpus  346
  12.5.8 Novel Topic Detection  350
  12.5.9 Topic Clustering  350
12.6 New Trends and Future Directions  352
References  353

13 Speech Summarization  357
Yang Liu and Dilek Hakkani-Tür
13.1 Task Description  357
  13.1.1 General Definition of Summarization  357
  13.1.2 Speech Summarization  359
  13.1.3 Applications  361
13.2 Challenges when Using Speech Input  362
  13.2.1 Automatic Speech Recognition Errors  363
  13.2.2 Speaker Turns  363
  13.2.3 Sentence Boundaries  363
  13.2.4 Disfluencies and Ungrammatical Utterances  364
  13.2.5 Other Style and Structural Information  365
13.3 Data Sets  366
  13.3.1 Broadcast News (BN)  367
  13.3.2 Lectures  368
  13.3.3 Multi-party Conversational Speech  369
  13.3.4 Voice Mail  371
13.4 Evaluation Metrics  371
  13.4.1 Recall, Precision, and F-measure  372
  13.4.2 ROUGE  372
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>13.4.3</td>
<td>The Pyramid Method</td>
<td>373</td>
</tr>
<tr>
<td>13.4.4</td>
<td>Weighted Precision</td>
<td>374</td>
</tr>
<tr>
<td>13.4.5</td>
<td>SumACCY and Weighted SumACCY</td>
<td>374</td>
</tr>
<tr>
<td>13.4.6</td>
<td>Human Evaluation</td>
<td>375</td>
</tr>
<tr>
<td>13.4.7</td>
<td>Issues and Discussions</td>
<td>375</td>
</tr>
<tr>
<td>13.5</td>
<td>General Approaches</td>
<td>375</td>
</tr>
<tr>
<td>13.5.1</td>
<td>Extractive Summarization: Unsupervised Methods</td>
<td>376</td>
</tr>
<tr>
<td>13.5.2</td>
<td>Extractive Summarization: Supervised Learning Methods</td>
<td>381</td>
</tr>
<tr>
<td>13.5.3</td>
<td>Moving Beyond Generic Extractive Summarization</td>
<td>385</td>
</tr>
<tr>
<td>13.5.4</td>
<td>Summary</td>
<td>386</td>
</tr>
<tr>
<td>13.6</td>
<td>More Discussions on Speech versus Text Summarization</td>
<td>386</td>
</tr>
<tr>
<td>13.6.1</td>
<td>Speech Recognition Errors</td>
<td>386</td>
</tr>
<tr>
<td>13.6.2</td>
<td>Sentence Segmentation</td>
<td>388</td>
</tr>
<tr>
<td>13.6.3</td>
<td>Disfluencies</td>
<td>389</td>
</tr>
<tr>
<td>13.6.4</td>
<td>Acoustic/Prosodic and Other Speech Features</td>
<td>390</td>
</tr>
<tr>
<td>13.7</td>
<td>Conclusions</td>
<td>391</td>
</tr>
<tr>
<td>References</td>
<td></td>
<td>392</td>
</tr>
</tbody>
</table>

14 Speech Analytics

I, Dan Melamed and Mazin Gilbert

14.1 Introduction | 397
14.2 System Architecture | 398
14.3 Speech Transcription | 401
14.4 Text Feature Extraction | 402
14.5 Acoustic Feature Extraction | 403
14.6 Relational Feature Extraction | 405
14.7 DBMS | 405
14.8 Media Server and Player | 408
14.9 Trend Analysis | 409
14.10 Alerting System | 413
14.11 Conclusion | 414
References | 415

15 Speech Retrieval

Ciprian Chelba, Timothy J. Hazen, Bhuvana Ramabhadran and Murat Saraçlar

15.1 Task Description | 417
15.1.1 Spoken Document Retrieval | 417
15.1.2 Spoken Utterance Retrieval | 418
15.1.3 Spoken Term Detection | 418
15.1.4 Browsing | 418
15.2 Applications | 418
15.2.1 Broadcast News | 419
15.2.2 Academic Lectures | 419
15.2.3 Sign Language Video | 419
15.2.4 Historical Interviews | 420
15.2.5 General Web Video | 420