

Contents

<i>Preface</i>	<i>page xi</i>
I Foundations	1
1 Introduction	3
1.1 Overview	4
1.2 History of Machine Translation	14
1.3 Applications	20
1.4 Available Resources	23
1.5 Summary	26
2 Words, Sentences, Corpora	33
2.1 Words	33
2.2 Sentences	45
2.3 Corpora	53
2.4 Summary	57
3 Probability Theory	63
3.1 Estimating Probability Distributions	63
3.2 Calculating Probability Distributions	67
3.3 Properties of Probability Distributions	71
3.4 Summary	75
II Core Methods	79
4 Word-Based Models	81
4.1 Machine Translation by Translating Words	81
4.2 Learning Lexical Translation Models	87
4.3 Ensuring Fluent Output	94
4.4 Higher IBM Models	96
4.5 Word Alignment	113
4.6 Summary	118

5	Phrase-Based Models	127
5.1	Standard Model	127
5.2	Learning a Phrase Translation Table	130
5.3	Extensions to the Translation Model	136
5.4	Extensions to the Reordering Model	142
5.5	EM Training of Phrase-Based Models	145
5.6	Summary	148
6	Decoding	155
6.1	Translation Process	156
6.2	Beam Search	158
6.3	Future Cost Estimation	167
6.4	Other Decoding Algorithms	172
6.5	Summary	176
7	Language Models	181
7.1	N-Gram Language Models	182
7.2	Count Smoothing	188
7.3	Interpolation and Back-off	196
7.4	Managing the Size of the Model	204
7.5	Summary	212
8	Evaluation	217
8.1	Manual Evaluation	218
8.2	Automatic Evaluation	222
8.3	Hypothesis Testing	232
8.4	Task-Oriented Evaluation	237
8.5	Summary	240
III	Advanced Topics	247
9	Discriminative Training	249
9.1	Finding Candidate Translations	250
9.2	Principles of Discriminative Methods	255
9.3	Parameter Tuning	263
9.4	Large-Scale Discriminative Training	272
9.5	Posterior Methods and System Combination	278
9.6	Summary	283
10	Integrating Linguistic Information	289
10.1	Transliteration	291
10.2	Morphology	296
10.3	Syntactic Restructuring	302

10.4 Syntactic Features	310
10.5 Factored Translation Models	314
10.6 Summary	320
11 Tree-Based Models	331
11.1 Synchronous Grammars	331
11.2 Learning Synchronous Grammars	337
11.3 Decoding by Parsing	346
11.4 Summary	363
<i>Bibliography</i>	371
<i>Author Index</i>	416
<i>Index</i>	427