

Contents

1	Introduction	1
1.1	Statistical Machine Translation	2
1.2	Phrase- and Syntax-Based SMT	4
1.3	Linguistic Challenges for Machine Translation	6
1.3.1	Lexical Challenge	6
1.3.2	Syntactic Challenge	7
1.3.3	Sentence-Level Semantic Challenge	7
1.3.4	Document-Level Semantic Challenge	7
1.4	Linguistically Motivated SMT	8
1.5	What This Book Is	9
1.6	Organization	10
2	BTG-Based SMT	15
2.1	Bracketing Transduction Grammar	16
2.2	A Unified Framework for BTG-Based SMT	17
2.2.1	Model	18
2.2.2	The $-LM$ Decoding Algorithm	19
2.2.3	Reordering	21
2.3	n -Gram Language Model Integration	22
2.4	Two Extensions to n -Gram Language Model	25
2.4.1	Backward Language Model	26
2.4.2	Trigger Language Model	29
2.5	Pruning	33
2.5.1	Dynamic Threshold Pruning	34
2.5.2	LM-Dependent Threshold Pruning	36
2.6	Comparison with Other SMT Formalisms	38
2.6.1	Comparison with Phrase-Based SMT	38
2.6.2	Comparison with Hierarchical Phrase-Based SMT	40
2.6.3	Comparison with Linguistically Syntax-Based SMT	40
2.6.4	Strengths and Weaknesses of BTG-Based SMT	41
2.7	Summary and Additional Readings	41

3	Syntactically Annotated Reordering	43
3.1	Reordering Example	45
3.2	Reordering Example Extraction Algorithm	45
3.2.1	AExtractor: Extracting Reordering Examples from Word Alignments	45
3.2.2	TExtractor: Extracting Reordering Examples from BTG-Style Trees	47
3.3	Boundary Word Reordering	49
3.3.1	Model	50
3.3.2	Analysis of Reordering Example Exaction Algorithms	51
3.4	Syntactically Annotated Reordering	53
3.4.1	Annotation Algorithm	53
3.4.2	Syntactically Annotated Reordering Model	56
3.4.3	Combining SAR and BWR	57
3.5	Syntax-Based Reordering Analysis: Method	58
3.5.1	Overview	58
3.5.2	Generating Word Alignments	59
3.5.3	Generating Target Spans and Orders	60
3.5.4	Generating Syntactic Reordering Patterns	61
3.6	Syntax-Based Reordering Analysis: Results	62
3.6.1	Syntactic Constituent Movement: Overview	62
3.6.2	Syntactic Constituent Movement in Reference Translations	63
3.6.3	Syntactic Constituent Movement in System Translations	65
3.6.4	Syntactic Reordering Patterns with Gaps	68
3.6.5	Challenges in Phrase Reordering and Suggestions	68
3.7	Summary and Additional Readings	69
4	Semantically Informed Reordering	71
4.1	Semantically Informed Argument Reordering	72
4.1.1	Model	72
4.1.2	Features	74
4.1.3	Training	75
4.2	Integration	76
4.3	Analysis	78
4.4	Summary and Additional Readings	79
5	Lexicalized Bracketing	81
5.1	Bracketable Segment and Its Boundaries	83
5.2	Learning Bracketable Segment Boundaries	84
5.3	Building Bracketable Segment Boundary Classifiers	86
5.3.1	Classifiers	86
5.3.2	Features	86

5.4	Penalty Model	87
5.5	Bracketing Strength Model	88
5.6	Integration.	89
5.7	Summary and Additional Readings.	91
6	Linguistically Motivated Bracketing	93
6.1	Syntax-Driven Bracketing	94
6.1.1	Learning Syntax-Driven Bracketing Instances	95
6.1.2	Model	96
6.1.3	Syntax-Driven Features	97
6.1.4	Integrating the SDB Model into SMT.	98
6.1.5	Comparing the SDB Model Against XP+	99
6.2	Semantically Informed Argument Bracketing.	101
6.2.1	Model	102
6.2.2	Features	102
6.2.3	Training	103
6.2.4	Integrating the Model into SMT	104
6.3	Summary and Additional Readings.	106
7	Translation Rule Selection with Document-Level Semantic Information.	107
7.1	Preliminaries	110
7.1.1	Hierarchical Phrase-Based SMT.	110
7.1.2	Topic Modeling.	111
7.2	Topic-Based Dissimilarity and Sensitivity Models	112
7.2.1	Topic Dissimilarity Model	112
7.2.2	Topic Sensitivity Model	113
7.3	Estimation.	114
7.3.1	Rule-Topic Distribution Estimation	116
7.3.2	Target-Side Rule-Topic Distribution Projection	116
7.4	Integration.	118
7.5	Discussion on Bilingual Topic Modeling.	121
7.6	Summary and Additional Readings.	121
8	Translation Error Detection with Linguistic Features.	125
8.1	Features	126
8.1.1	Lexical Features	127
8.1.2	Syntactic Features	127
8.1.3	Word Posterior Probability Features	128
8.2	Error Detection with a Maximum Entropy Model.	129
8.3	Data Collection	130
8.4	Evaluation Metrics	131
8.5	Architecture of Translation Error Detection System	131
8.6	Summary and Additional Readings.	133

9 Closing Remarks	137
9.1 Linguistically Motivated SMT: Review	137
9.2 Linguistically Motivated SMT: Future Directions	138
References.	141
Index	151

Linguistically Motivated Statistical Machine Translation
Models and Algorithms

Xiong, D.; Zhang, M.

2015, XII, 152 p. 52 illus., Hardcover

ISBN: 978-981-287-355-2