

Contents

1	Data Acquisition and Linguistic Resources	1
1.1	Introduction	1
1.2	Data Collection, Distribution, and Management	2
1.3	Human Annotation	14
1.4	Automatic Annotation	64
2	Machine Translation from Text	133
2.1	Introduction	133
2.2	Segmentation, Tokenization and Preprocessing	135
2.3	Word Alignment	164
2.4	Translation Models	183
2.5	Language Modeling for SMT	252
2.6	Search and Complexity	271
2.7	Adaptation and Data Selection	297
2.8	System Combination	324
3	Machine Translation from Speech	399
3.1	Introduction	399
3.2	Front End Features	401
3.3	Improved Speech Acoustic Models	428
3.4	Language Models	460
3.5	Language-Specific Models and Systems: Mandarin	485
3.6	Language-Specific Models and Systems: Arabic	520
3.7	Integration of Speech Recognition and Translation	569
4	Distillation	617
4.1	Introduction	617
4.2	Template-Based Query Development	618
4.3	Architecture and Implementation of a Distillation System	623
4.4	Enabling Technology Breakthroughs to Improve Distillation Capabilities	636
4.5	Distillation in an Integrated GALE system	690
4.6	Evaluating Distillation Technology	716
5	Machine Translation Evaluation and Optimization	745
5.1	Introduction	745
5.2	Automatic and Semi-Automatic Measures	758
5.3	Tasks and Human-in-the-Loop Measures	768

5.4	GALE Machine Translation Metrology: Definition, Implementation, and Calculation	783
5.5	Use of Evaluation for Optimization	812
5.6	Searching for Better Automatic MT Metrics	818
6	Operational Engines	845
6.1	Introduction	845
6.2	Implementation of Operational Engines	846
6.3	Evaluation of Operational Engines	905
	Concluding Remarks	933

Handbook of Natural Language Processing and
Machine Translation

DARPA Global Autonomous Language Exploitation

Olive, J.; Christianson, C.; McCary, J. (Eds.)

2011, XXVI, 936 p., Hardcover

ISBN: 978-1-4419-7712-0